SVEDBERG User Manual

SVEDBERG User Manual

Version 7.0.4
September 5, 2013

SVEDBERG User Manual

Table of Contents

1. Getting Started 1
   1.1. Program overview and organization 1-2
   1.2. How to use this Help file 2-3
   1.3. Features for SEDFIT users 3
2. Tutorials 4
   2.1. Tutorial: Single species fit 4-34
   2.2. Tutorial: Multi-species fits 34-55
   2.3. Tutorial: Using a c(s) distribution with SVEDBERG 56-77
   2.4. Tutorial: Simulating experiments with the Claverie simulator 77-80
3. Program elements, controls, and features 81
   3.1. User interface 81
      3.1.1. Main menus 81
         3.1.1.1. File menu 81-82
            3.1.1.1.1. Open saved analysis... 82-83
            3.1.1.1.2. Save analysis 83
            3.1.1.1.3. Save analysis as... 83
            3.1.1.1.4. Clone analysis to new window 83
         3.1.1.2. Edit menu 83-85
         3.1.1.3. View menu 85
         3.1.1.4. Tools menu 85
            3.1.1.4.1. Mass/diffusion calculator 85-87
            3.1.1.4.2. F statistics calculator 87-89
            3.1.1.4.3. Options dialog 89-93
         3.1.1.5. Windows menu 93-94
         3.1.1.6. Help menu 94-96
      3.1.2. Toolbar 96
      3.1.3. Analysis window 96-98
         3.1.3.1. Next and Back buttons 98
         3.1.3.2. Log page 98-100

3.1.3.3. User prompt area 100
3.1.4. Navigation tree 100-102
3.2. Analysis pages 102
3.2.1. Load raw scans page 102-104
3.2.1.1. Load raw scans - Preselect data type and cell check box 104
3.2.1.2. Load raw scans - Load new scan files button 104-105
3.2.1.3. Load raw scans - Replace scans for this sample button 105
3.2.1.4. Load raw scans - Add more scans for this sample button 105
3.2.1.5. Load raw scans - Delete selected scans button 105-106
3.2.1.6. Load raw scans - Scans currently loaded in memory list box 106
3.2.1.7. Load raw scans - Data loading options 106-107
3.2.1.8. Baseline scan usage 107-109
3.2.1.9. Load raw scans - Sample description 109
3.2.1.10. Load raw scans - Run Claverie simulation to generate data button 109-110
3.2.2. Set meniscus & data region page 110-112
3.2.2.1. Graph display options tab 112-113
3.2.2.2. Systematic noise corrections options tab 113-114
3.2.2.2.1. Set meniscus & data region - Jitter removal options 114-115
3.2.2.2.2. Set meniscus & data region - Fringe jump removal options 115-116
3.2.2.3. Set meniscus & data region - Meniscus position 116-117
3.2.2.4. Set meniscus & data region - Good data begin X cm from meniscus 117
3.2.2.5. Good data end at radius = 117-118
3.2.2.6. Set meniscus & data region - Zoom meniscus button 118
3.2.2.7. Cell base position 118
3.2.2.8. Limit maximum buildup at cell base 118-119
3.2.3. Select fitting model page 119-121
3.2.3.1. Select fitting model - Multi-species options 121-122
3.2.3.2. Select fitting model - Import from SEDFIT c(s) button 122-123
3.2.3.3. Select fitting model - Fitting function options 123-124
3.2.3.4. Data options 124-131
3.2.3.5. Baseline fitting options 131-132

3.2.3.6. Select fitting model - Fit diffusion or mass 133
3.2.3.7. Select fitting model - Constrain M or s ratios among species 133-136
3.2.3.8. Select fitting model - Molar mass units 136-137
3.2.3.9. Select fitting model - Density and partial specific volumes 137-138
3.2.3.10. Convert to s(20,w) 138-140
3.2.4. Set/alter fitting parameters page 140-142
3.2.4.1. Parameter value text boxes 142
3.2.4.2. Fit control check boxes 142-143
3.2.4.3. Constrained parameters 143
3.2.4.4. Convergence control 143-144
3.2.4.5. Adjust parameter upper/lower limits button 144-145
3.2.4.6. Slope correction fitting parameter 145-146
3.2.4.7. Cancel changes button 146
3.2.5. Perform least-squares fit page 146-148
3.2.5.1. Perform fit - Do fit button 148
3.2.5.2. Perform fit - 1 iteration button 148-149
3.2.5.3. Perform fit - Simulate button 149
3.2.5.4. Perform fit - Cancel button 149
3.2.5.5. Perform fit - Status 149
3.2.5.6. Fitting parameters monitor tab 149-151
3.2.5.6.1. Increment reduction factor 151-152
3.2.6. Report fit results page 152-154
3.2.6.1. Report fit results - Fitted parameters 154-155
3.2.6.2. Report fit results - Fixed, constrained, & implicit parameters 155
3.2.6.3. Report fit results - View parameter cross-correlations button 155-156
3.2.6.4. Report fit results - Compute confidence intervals button 156-159
3.2.6.5. Report fit results - Print reports 159-160
3.2.6.6. Report fit results - Fit statistics 160-161
3.2.6.7. Report fit results - High cross-correlation warning 161
3.2.6.8. Report fit results - Graph bootstrap or Monte Carlo results 161-164
3.2.7. Graph fit results page 164-181

4. How to... 182
4.1. How to avoid common mistakes 182-183
4.2. How to optimize your experiments 183
4.2.1. How (and why) to equilibrate the rotor temperature 183-184
4.3. How to work with graphs 184-185
4.3.1. How to print graphs 185
4.3.2. How to customize graphs 185-188
4.3.3. How to export graph images or data to other programs 188-189
4.3.4. How to zoom with the mouse 189
4.4. How to remove bad data points from scan files 189-190
4.5. How to identify and remove rogue scans 190-194
4.6. How to minimize the effects of a sloping plateau 194-195
5. Frequently-asked questions 196
5.1. About fitting 196
5.1.1. Should I fit the meniscus position? 196-197
5.1.2. Why should I limit the buildup at the cell base? 197
5.1.3. Should I fit the cell base position? 197-198
5.1.4. Which method of systematic noise removal should I use? 198-201
5.1.5. What is a 'sloping plateau'? 201-202
5.1.6. What does high parameter cross-correlation mean? 202
5.1.7. What constitutes a good fit? 202-203
5.2. Does SVEDBERG utilize multi-core processors? 203
5.3. How should I cite this program? 203
6. Reference information 204
6.1. Symbols used 204
6.2. About using radial intensity scans 204-205
6.3. About the fitting algorithm 205-206
6.4. About the meniscus wizard 206
6.5. About the jitter wizard 206
6.6. About the bootstrap method 206-207
6.7. About the Monte Carlo method 207-208

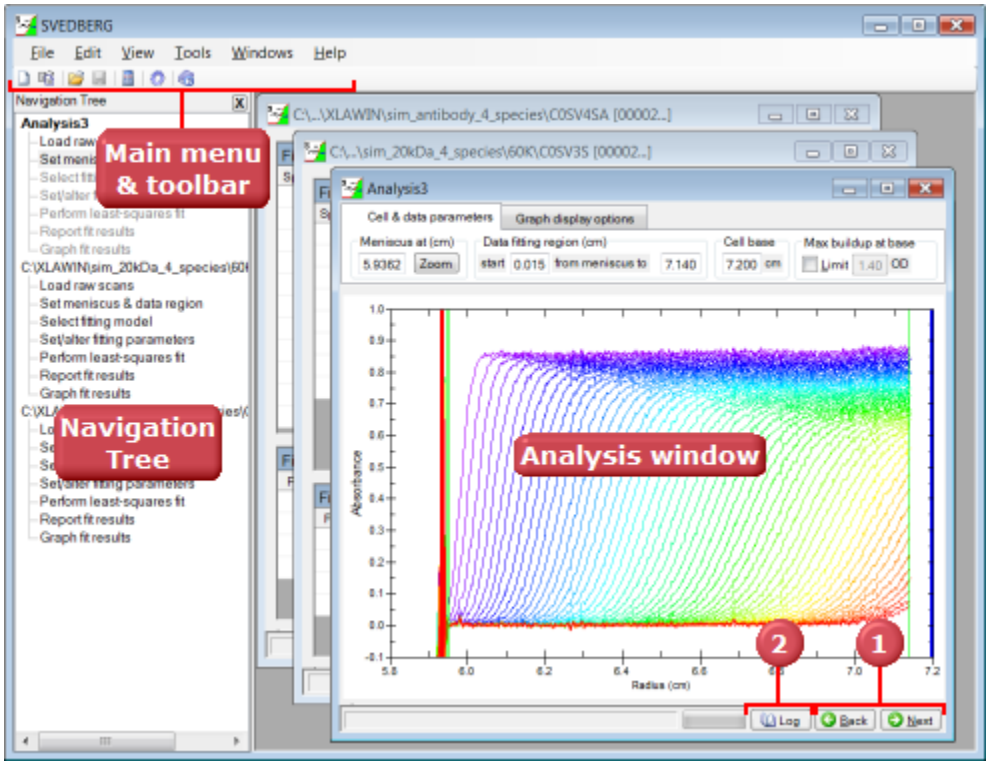
7. Bibliography 209
7.1. reference 1 - Philo (1997) 209
7.2. reference 2 - Johnson & Faunt (1992) 209
7.3. reference 3 - Stafford (1997) 209
7.4. reference 4 - Durschlag (1986) 209
7.5. reference 5 - Behlke & Ristau (2002) 209
7.6. reference 6 - Balbo et al. (2009) 209
7.7. reference 7 - Stafford (1994) 210
7.8. reference 8 - Schuck & Demeler 1999 210
7.9. reference 9 - Efron & Tibshirani (1986) 210
7.10. reference 10 - Garcia de la Torre & Bloomfield (1981) 210
7.11. reference 11 - Johnson & Straume (1994) 210
7.12. reference 12 - Philo (2011) 210
7.13. reference 13- Arthur et al. (2009) 210-211

1 Getting Started

1.1 Program overview and organization

Below are pictures of the full program screen to illustrate various parts of the program user interface, followed text explaining more about the program design and operation. (See **How to use this Help file (Section 1.2)** for information about *expandable sections* and other features of this Help file).

The **See Also section** (at the bottom, or use the button at upper left) will lead you to tutorials and other information to help get you started.



Main menu & toolbar SVEDBERG has a main program menu with the standard File, View, Window and other main headings you find in many Windows programs. Analysis documents can be created, opened and saved using the familiar New, Open, Save, and Save As... commands under the File menu, or using the familiar icons on the Toolbar.

Analysis window An analysis document within SVEDBERG is contained within an **Analysis window (Section 3.1.3)**. In the example above, three analysis windows are open. Each analysis window concerns the analysis of one sample. In each **analysis window (Section 3.1.3)** you can work through the analysis in a step-by-step manner, and different controls and/or graphs are shown as different pages within the window at different steps. The analysis window in the foreground is at the second step, the **Set meniscus & data region page (Section 3.2.2)**. The two analysis pages in the background are at the **Report fit results page (Section 3.2.6)**.

Navigating within and between analysis pages

1 Movement among the steps (pages) within one analysis window can be accomplished using either the **Next button** and **Back button** at the lower right, or by using the **Navigation tree (Section 3.1.4)** (see below), or by right-clicking on the analysis window with the mouse and using the Go to step popup menu.

Navigation Tree SVEDBERG also provides an Explorer-style **Navigation tree (Section 3.1.4)** along the left side. This navigation tree provides an easy mechanism to move both between different analysis windows, and also to specific pages (steps) within each analysis window.

The Log page

2 A **Log page (Section 3.1.3.2)** keeps a permanent record of every analysis step and choice, and any intermediate analysis results (including graphs summarizing each fit). This log can be viewed at any time by pressing the Log button, and can then be printed, or pasted into your electronic lab notebook. This **Log page (Section 3.1.3.2)** is not part of the sequence of analysis pages, and those pages are hidden while the Log is displayed.

1.2 How to use this Help file

Within certain topics you will see *expandable sections* like the one below that when clicked expand to provide more details about that subject.

▸ **Getting context-sensitive Help from within the program**

As is true for many Windows® programs, pushing the F1 key at any point in the program will bring up a Help screen relevant to that context (page within the analysis window, dialog box, or control).

Remember too that virtually every button or other control will display brief "tool tip" help if you hover over that control with your mouse.

For pages with multiple expandable links you may also see a **Show All/Hide All** link like the one at the right below that expands or contracts all such material within that page.

Special text sections

Tips about easier or better ways to accomplish tasks will be indicated by a colored block with the **idea icon**, as shown below:

TIP: To best way to do X is to use the Y button and then enter...

Cautions about potential mistakes or pitfalls will be indicated by a colored block with a **caution icon**, as shown below:

CAUTION: Be sure the rotor temperature is equilibrated before starting your run---even the best software cannot overcome bad data!

Material that provides **notes, background information, or details** about program features or operations will be indicated by a colored block with the **teacher icon**, as shown below:

The intellectually curious user will want to know that...

Such information may also be found within an expandable text block in some cases.

Short blocks of text about **advanced subjects** or features primarily of interest to **expert users** will be indicated by a colored block with the **Einstein icon**, as shown below:

This is text concerning a feature that is probably only of interest to expert users.

Long blocks of text about advanced subjects or features primarily of interest to experts may be indicated simply by the Einstein icon.

Send Comments link

At the upper right corner of every topic is a link that allows you to send an e-mail to the author if you find errors in the Help file, if you have requests for new topics or tutorials, or if you have other comments.

See Also links

A **See Also** section like that below that leads you to related topics and further information will be found at the bottom of many topics. The **See Also** link symbol, , at the upper left corner of the topic page, takes you directly to the **See Also** list.

1.3 Features for SEDFIT users

SVEDBERG includes some features intended to make things easier for those of you who also use SEDFIT. These include:

- If you want to load the same scans you analyzed using SEDFIT, during file loading simply specify the SEDFIT list file (for example list.ra1) rather than the individual data files. SVEDBERG will then automatically check for a SEDFIT parameters file, and if it exists will ask you whether you want to use the same data fitting region, meniscus position, *etc*. that you used in SEDFIT.
- If you have generated and saved a *c(s)* distribution for this sample with SEDFIT, you can use that distribution to create a corresponding discrete species model in SVEDBERG. The *c(s)* distribution is automatically integrated and the number of peaks is determined, and then the concentrations, sedimentation coefficients, and apparent masses of those peaks are automatically entered as initial guesses for the SVEDBERG model. Just as in *c(s)*, the masses of different species will be constrained using an assumption that all species have the same hydrodynamic shape ($f/f0$ ratio). This can allow you to get true confidence intervals for the properties of those species, and also allows you to then explore removing the constant $f/f0$ ratio constraint(s). See **Tutorial: Using a c(s) distribution with SVEDBERG (Section 2.3)**.
- As in SEDFIT, the keyboard shortcut CTRL-L will load new scan files (this shortcut is equivalent to pushing the **Load new scan files button (Section 3.2.1.2)**.
- SVEDBERG recognizes that SEDFIT creates temporary files with file extensions that suggest they are scan files (files with names such as ~tmppars.ra1, list.ra1, and sdist.ra1), and will ignore them if you happen to select them as scans to be loaded. This allows you to quickly and easily choose to load all available scans by clicking any file and using the Windows shortcut key CTRL-A to select all of them.

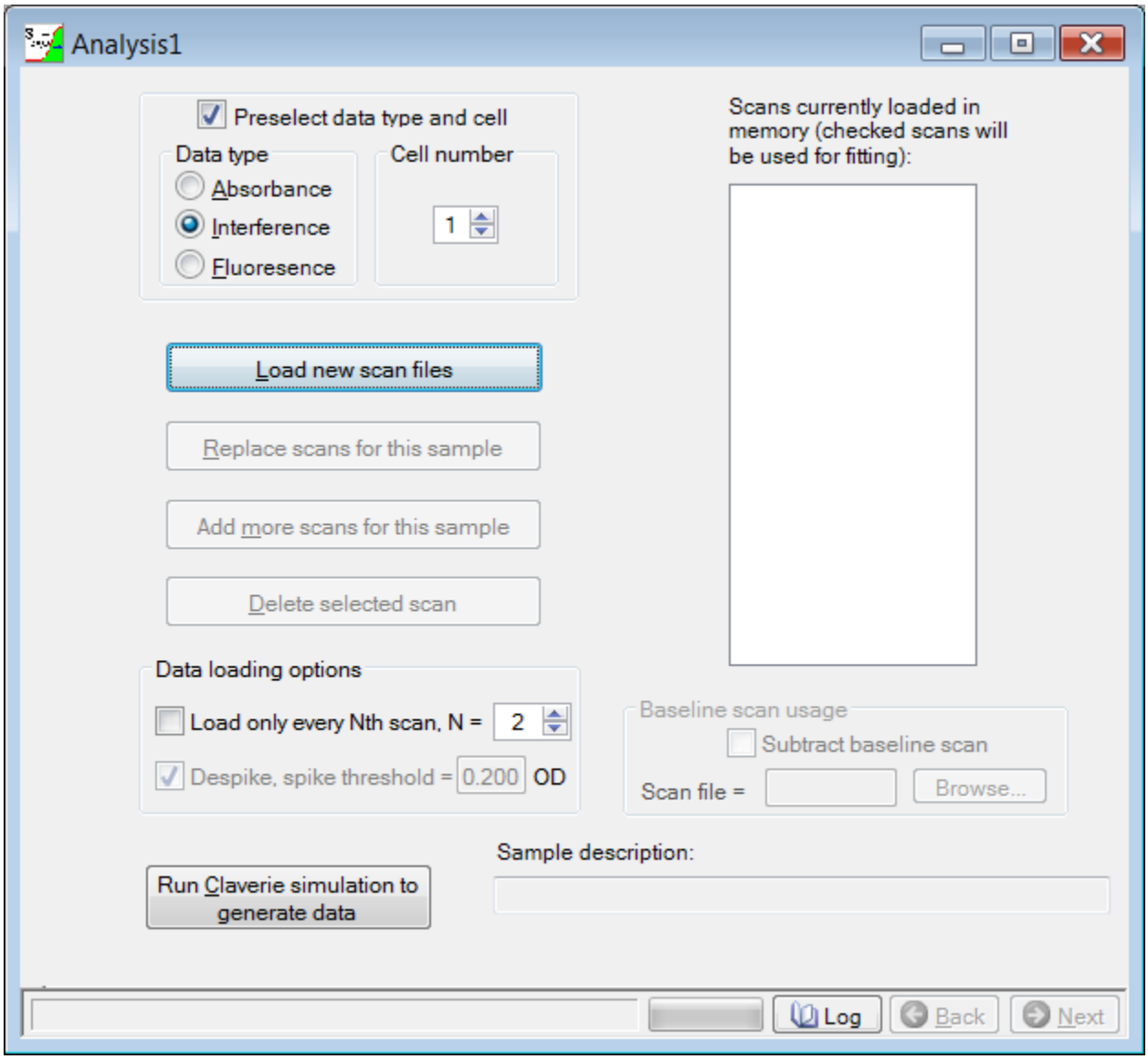
2 Tutorials

2.1 Tutorial: Single species fit

Below is a detailed and general procedure for loading scans and fitting them to a single-species model. It looks more complicated than it is---this takes only a few minutes.

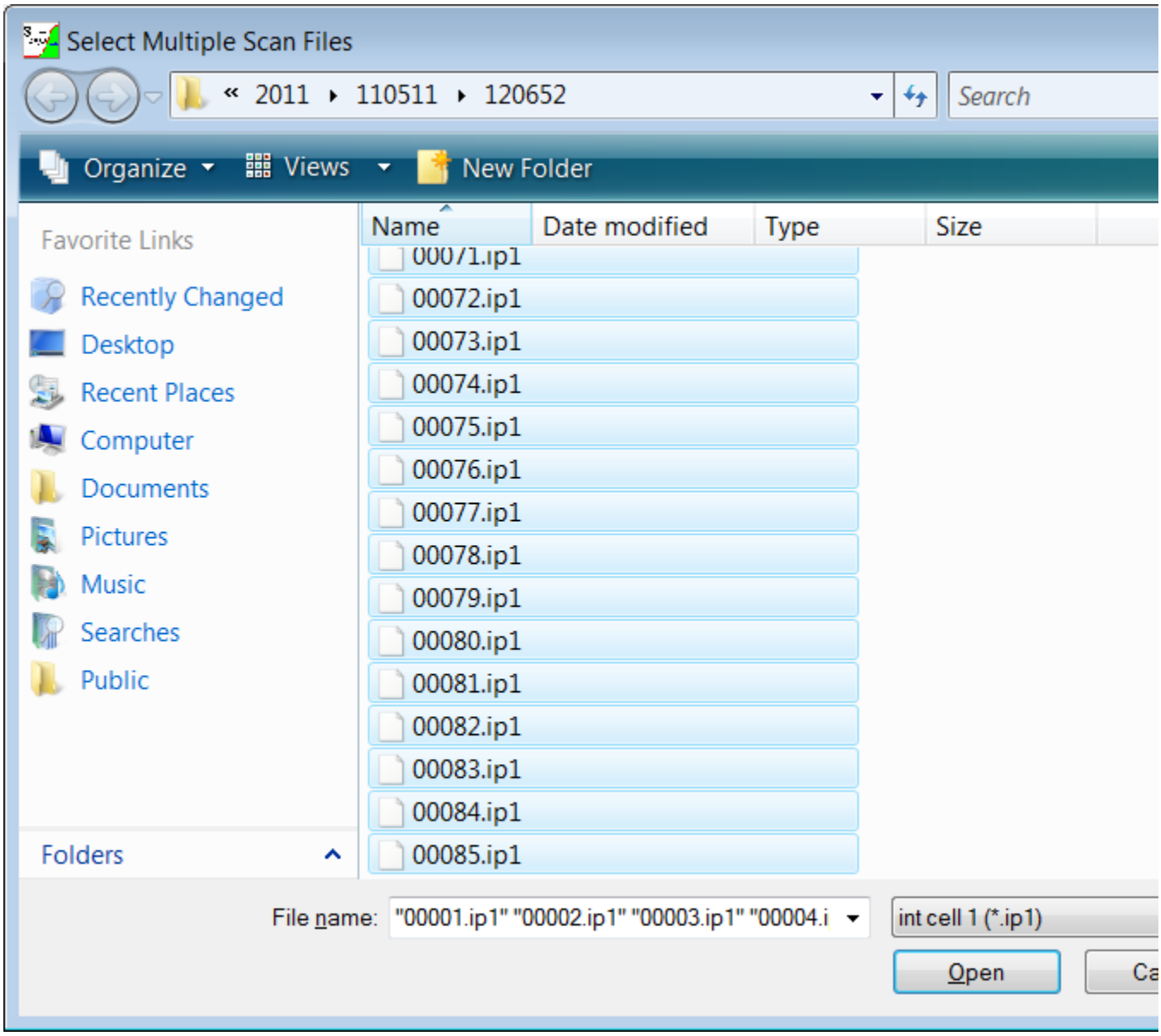
1. Load the scans to be analyzed

(1) Begin on the **Load raw scans page** **(C:\Users\Public\Documents\Visual Studio Projects\Svedberg Help Source\{B80B8F9E-AE00-4422-B154-B929EC55899B})** of a new or existing **analysis window** **(C:\Users\Public\Documents\Visual Studio Projects\Svedberg Help Source\{9FF3D5BE-B726-407D-92FA-C14A83E56F99})**. Usually it is a good idea to **(C:\Users\Public\Documents\Visual Studio Projects\Svedberg Help Source\{A7B32D92-2FF9-4DF8-B7B1-B40DEE5398B8})****preselect the data type and cell number** **(C:\Users\Public\Documents\Visual Studio Projects\Svedberg Help Source\{A7B32D92-2FF9-4DF8-B7B1-B40DEE5398B8})** for the data you wish to load. In this example, interference data from cell 1 are pre-selected.



If you have collected hundreds of interference scans at a rapid rate, at least initially you might want to check the **Load only every Nth scan** option at this point and set the value of N such that you will be fitting 100 scans or less. In this particular experiment both absorbance and interference scans were collected, so there are only 85 of each type.

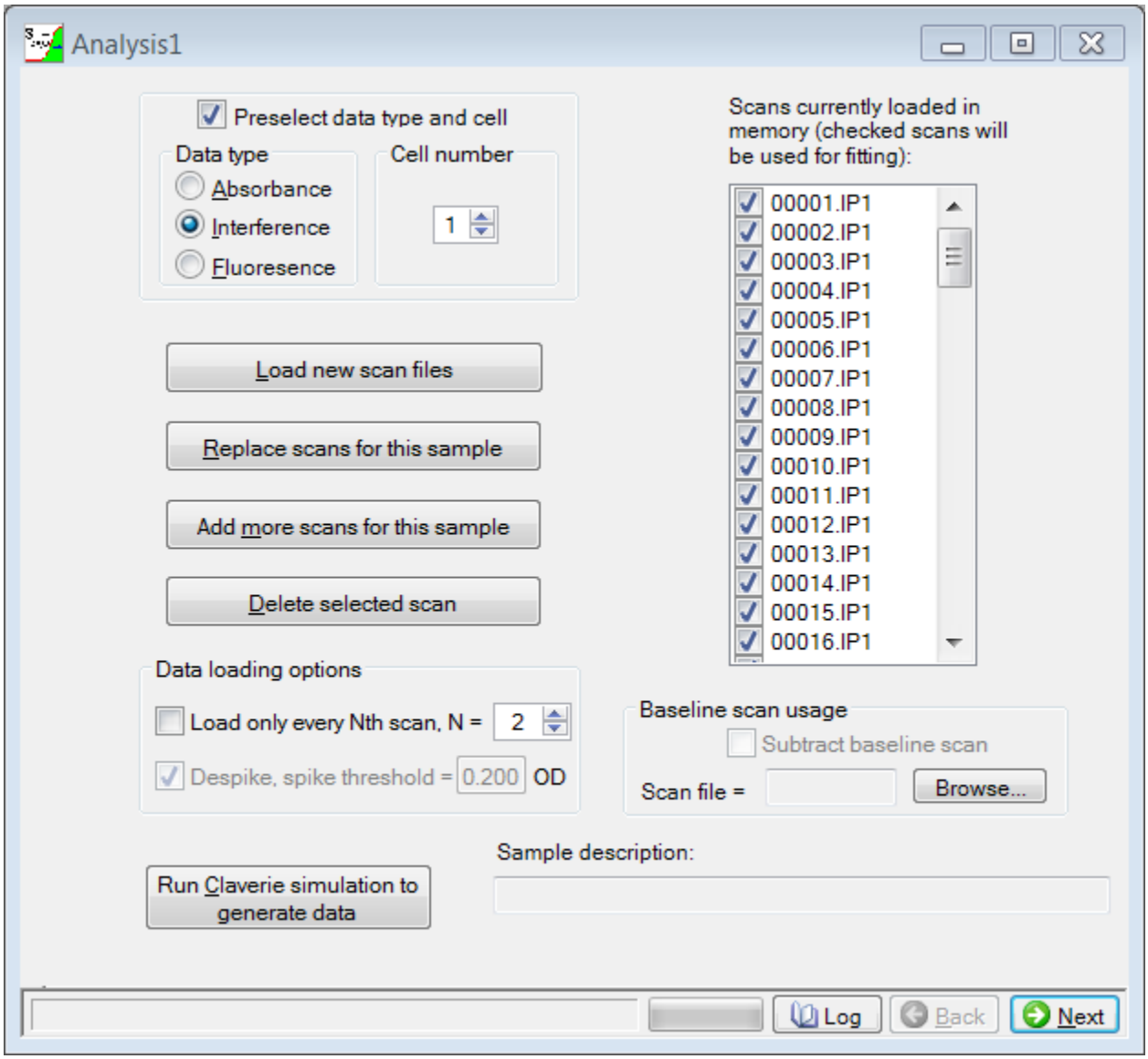
(2) Then push the **Load new scan files** **(C:\Users\Public\Documents\Visual Studio Projects\Svedberg Help Source\{06352269-002D-4939-A2B0-2E532D3353B4})** button to bring up a multiple File Open dialog box and select the scans you wish to load.



If you think all the scans are appropriate for fitting then a useful trick is to click on any one scan and then enter the standard Windows **CTRL-A** shortcut to select all the scans.

TIP: You don't want to load a bunch of scans at the end of the run where the cell is completely empty. Such scans don't really contribute any needed information, and including them will just distort the statistics for the fit.

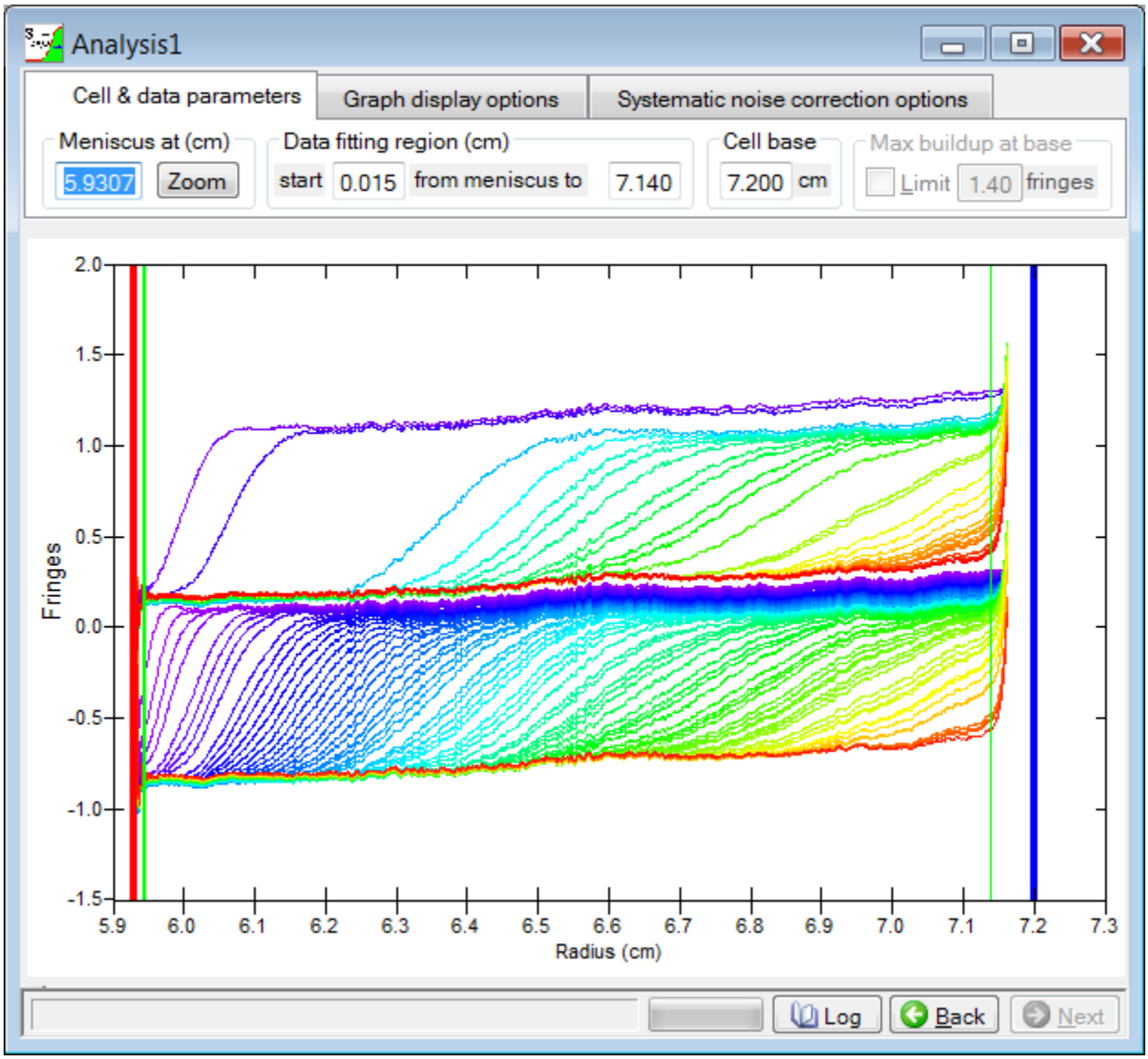
(3) Once the scans are selected click **Ok** in the File Open dialog to load these scans. Loading could require 30 seconds or more depending on the number of scans (and possibly network speed if you are accessing through a network). The progress bar will advance as the files load.



At this point all the scan names will show in the **Scans currently loaded in memory** **(C:\Users\Public\Documents\Visual Studio Projects\Svedberg Help Source\{6929048D-9483-421B-9B72-6CF14F1583C6})** list box, and your description of that sample will appear near the bottom of the page in the **sample description area** **(C:\Users\Public\Documents\Visual Studio Projects\Svedberg Help Source\{17E5DA8E-8F28-4E60-8DA7-C5738DD49027})**. The check marks next to each scan name indicate that these scans will be used during fitting---if ever needed, a bad scan can be removed simply by un-checking this box.

(4) Push the **Next button** **(C:\Users\Public\Documents\Visual Studio Projects\Svedberg Help Source\{46451615-072B-4FCC-B31F-82953CB24396})** to move to the **(C:\Users\Public\Documents\Visual Studio Projects\Svedberg Help Source\{14FCF551-188C-4C48-9E48-0B1B6957B1E6})Set meniscus & data region page** **(C:\Users\Public\Documents\Visual Studio Projects\Svedberg Help Source\{14FCF551-188C-4C48-9E48-0B1B6957B1E6})** and see a graph overlaying all of the scans you just loaded.

- 2. Set **the meniscus position and the data region for fitting**

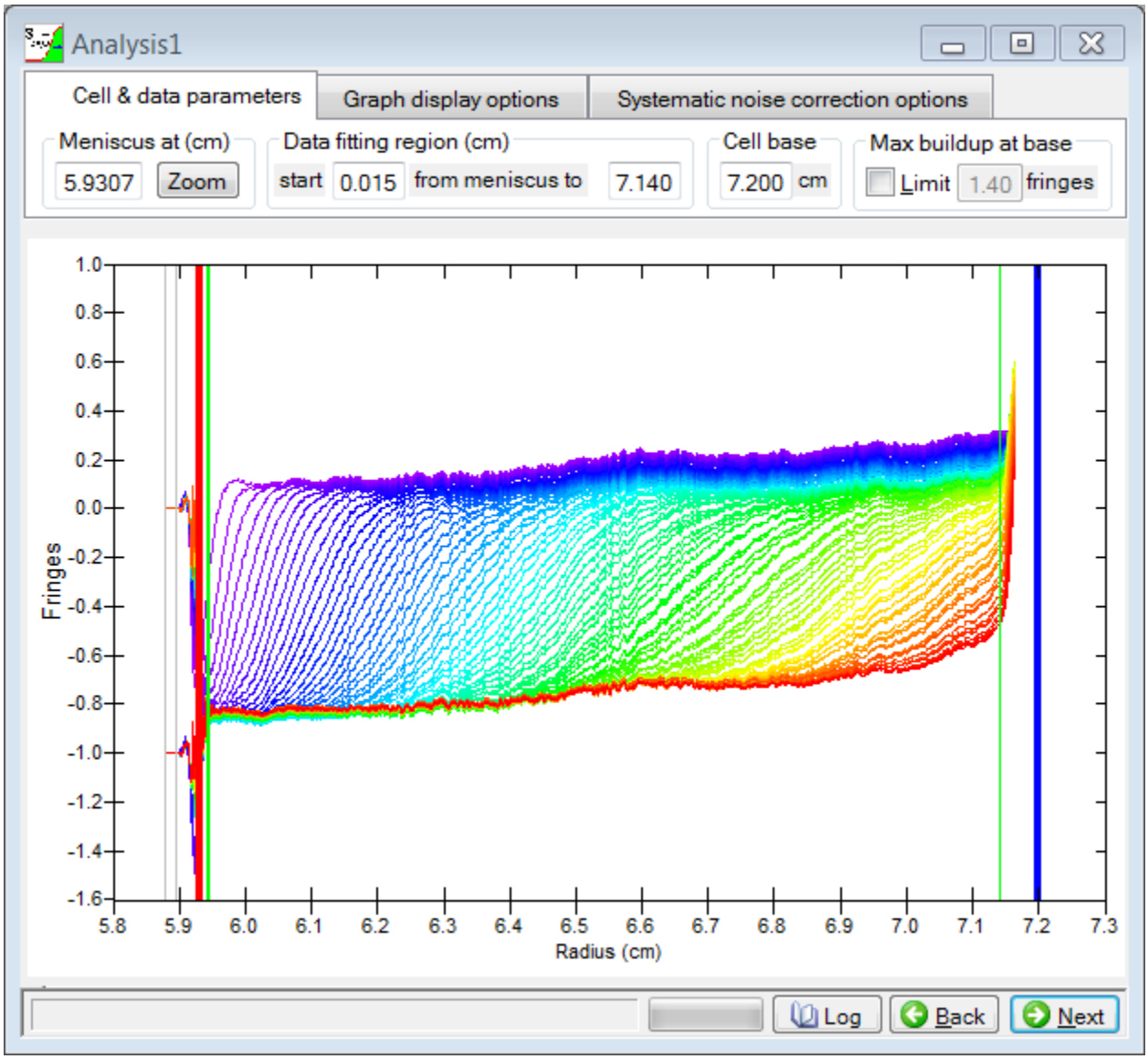


(1) When you first enter the the **Set meniscus & data region page (Section 3.2.2)** all the scans are graphed and the program's **meniscus wizard (Section 6.4)** will attempt to locate the meniscus position and will place the red vertical line at that position, as shown above. That position will be filled into the meniscus text box at the upper left, and that text box is automatically given the focus.

- At this point the *fringe jumps* in these scans make it somewhat difficult to see the boundary movement. Those will automatically be removed as soon as a meniscus position has been set.
- Note that later you may choose to treat the meniscus position as a fitting parameter, in which case the position set on this page just serves as the starting value (the initial guess) for the fit. See also **Should I fit the meniscus position? (Section 5.1.1)**

(2) If you are satisfied with the wizard's choice of meniscus position then all you need to do is push the **Enter** key on the keyboard to accept that choice. Otherwise you can click on the red meniscus marker and drag it to the position you want, or simply type a value into the text box.

- To set the meniscus more precisely you may wish to zoom the graph around the meniscus position. You can do that manually by dragging the mouse over the region on the graph (left-click and drag).



(3) In addition to setting the meniscus position, you also need to decide which portions of the scan will actually be included in the fit. That region falls between the two heavy green vertical lines shown on the graph.

- Usually the data very close to the meniscus are distorted by optical artifacts and thus should be excluded. By default the program excludes a region extending 0.015 cm to the right of the meniscus (the **start from** value shown above the graph). This value can be altered either by typing a new value in the text box or by clicking on the green vertical line on the left side and dragging it to the desired position.
- It may also be a necessary to adjust the fitting limit at the right, near the base of the cell, especially if the data in that region are being heavily influenced by the buildup of solutes at the base of the cell. In this experiment the data collection was stopped at 7.163 cm, and the fitting limit was set by default at 7.140 cm. With that setting there appears to be only a small influence of the buildup on the data being fitted. Thus this position is probably good if we don't want a strong influence of region near the cell base (but therefore possibly bad with respect to trying to fit for the exact cell base position). One could pull the green line to the left to about ~7.05 cm and thereby essentially remove the importance of exactly what is happening at the base of the cell, or move it all the way to the right to maximize that influence. When it is possible to do so it is usually a good idea to avoid having the buildup region exert a substantial influence on the fit because that region is subject to optical artifacts, and because the fitting models cannot accurately model all the things that happen there.

(4) It might also be necessary to adjust the position of the base of the cell, which is listed at the top and marked on the graph by a heavy blue vertical line. The default position is 7.20 cm, as shown here.

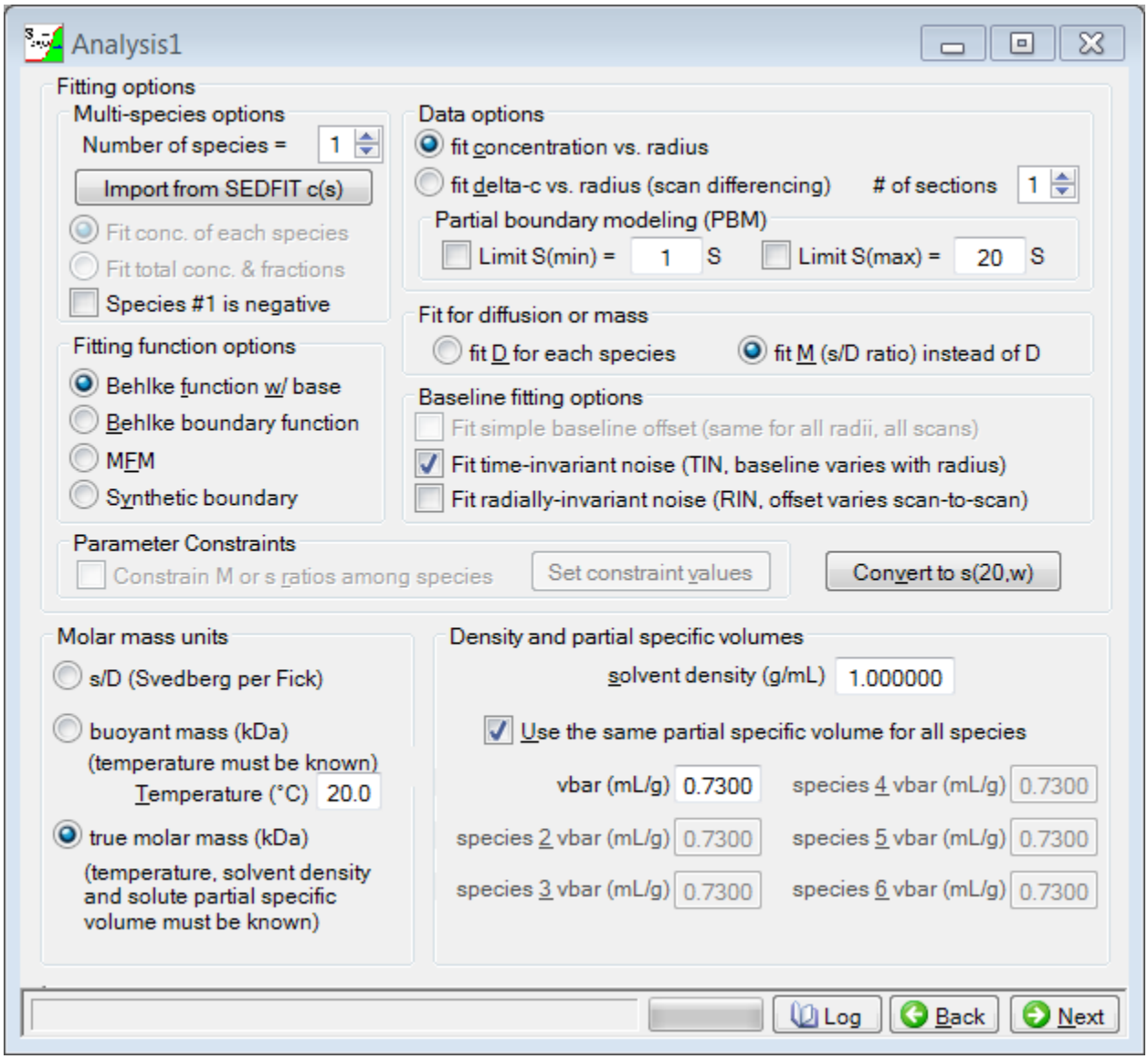
- At one extreme, certain samples show no buildup near the base of the cell because the macromolecules are precipitating there, and when this is true the exact location of the cell base is irrelevant.
- At the other extreme, when the macromolecule has a relatively low mass (or the rotor speed was too low) the data over large portions of the cell could be heavily influenced by the buildup. In this case the exact location of the cell base is a critical parameter and will probably need to be treated as a fitting parameter. For more discussion, see **Should I fit the cell base position? (Section 5.1.3)**

(5) This analysis page also includes an advanced option for controlling how much the details of what is happening at the base of the cell will influence the fit. It is possible to set an explicit limit on the maximum allowed buildup (in signal units), and then to adjust the right-hand end of the fitting region for each scan to cut off the data (if needed) to stay within that limit. This option is particularly useful for absorbance scans, since the buildup often reaches OD values well above the linear range. See **Why should I limit the buildup at the cell base? (Section 5.1.2)** for more details about why and when this is useful.

(6) Other advanced options can be found by clicking on either the **Graph display options tab (Section 3.2.2.1)** or the **Set meniscus & data region - Systematic noise corrections options tab (Section 3.2.2.2)** at the top of the page, but usually these are not needed.

(7) Push the **Next button (Section 3.1.3.1)** to move to the **Select fitting model page (Section 3.2.3)**.

3. Select a fitting model; enter vbar and solvent density


Analysis1
Fitting options
Multi-species options
Number of species = 1
Import from SEDFIT c(s)
Fit conc. of each species
Fit total conc. & fractions
Species #1 is negative
Fitting function options
Behlke function w/ base
Behlke boundary function
MFM
Synthetic boundary
Data options
fit concentration vs. radius
fit delta-c vs. radius (scan differencing)
# of sections 1
Partial boundary modeling (PBM)
Limit S(min) = 1 S
Limit S(max) = 20 S
Fit for diffusion or mass
fit D for each species
fit M (s/D ratio) instead of D
Baseline fitting options
Fit simple baseline offset (same for all radii, all scans)
Fit time-invariant noise (TIN, baseline varies with radius)
Fit radially-invariant noise (RIN, offset varies scan-to-scan)
Parameter Constraints
Constrain M or s ratios among species
Set constraint values
Convert to s(20,w)
Molar mass units
s/D (Svedberg per Fick)
buoyant mass (kDa)
(temperature must be known)
Temperature (°C) 20.0
true molar mass (kDa)
(temperature, solvent density and solute partial specific volume must be known)
Density and partial specific volumes
solvent density (g/mL) 1.000000
Use the same partial specific volume for all species
vbar (mL/g) 0.7300
species 4 vbar (mL/g) 0.7300
species 2 vbar (mL/g) 0.7300
species 5 vbar (mL/g) 0.7300
species 3 vbar (mL/g) 0.7300
species 6 vbar (mL/g) 0.7300
Log
Back
Next

Although there are many possible options on this **Select fitting model page (Section 3.2.3)**, many are only needed in special situations, and often the program default settings will suffice (at least for an initial fit). **The two things on this page that you will likely always want to change (or at least verify) are the solvent density and partial specific volume entries on the lower right portion of the page.** For this experiment the exact values are not known so we will leave them at nominal values (vbar = 0.73 mL/g, ρ = 1.00 g/mL).

Below some of the other options which *might* need to be altered are briefly discussed (but more detail is available in the specific Help page for each of the sections of this page).

(1) Each new analysis document initially assumes a single species model. Since that is what we want we do not need to adjust the number of species using the up/down spinner in the upper left corner.

(2) Usually by default the fitting function will include the buildup of solutes at the cell base, and that is appropriate for this experiment. However if no buildup is present in the data being fitting (which may be true for solutes of very high molar mass, or those which precipitate at the cell base, or if the fitting region terminates well to the left of cell base) then the buildup at the base should not be included in the fitting model. In such situations you should move to the **Fitting function options (Section 3.2.3.3)** group box and instead select the 'Behlke boundary function' option, which will omit calculating the buildup at the base. That option also speeds up the computations.

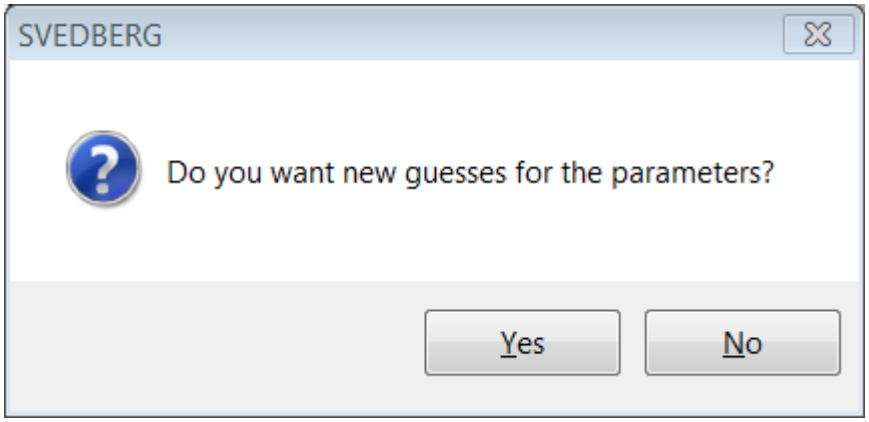
(3) This program provides multiple approaches to removing systematic noise from the scans, which are governed by settings in the **Data options (Section 3.2.3.4)** and **Baseline fitting options (Section 3.2.3.5)** sections. For a discussion of benefits/drawbacks of the different approaches see **Which method of systematic noise removal should I use? (Section 5.1.4)** For these interference scans we will remove the baseline noise or TIN by explicitly including it in the fitting model (as indicated by the check box within **Baseline fitting options (Section 3.2.3.5)**), but remove the *jitter* or RIN by aligning the fringes in the air-air space (as was automatically done previously on the **Set meniscus & data region page (Section 3.2.2)**.

(4) The initial program default is to characterize each species by it molar mass rather than its diffusion coefficient, as shown by the radio buttons in the **Fit diffusion or mass (Section 3.2.3.6)** section. In this example we will leave this unchanged.

We are ready to continue so push the **Next button (Section 3.1.3.1)** to move to the **Set/alter fitting parameters page (Section 3.2.4)**.

- **4. Confirm the initial guesses for the fitting parameters and which parameters should be fitted**

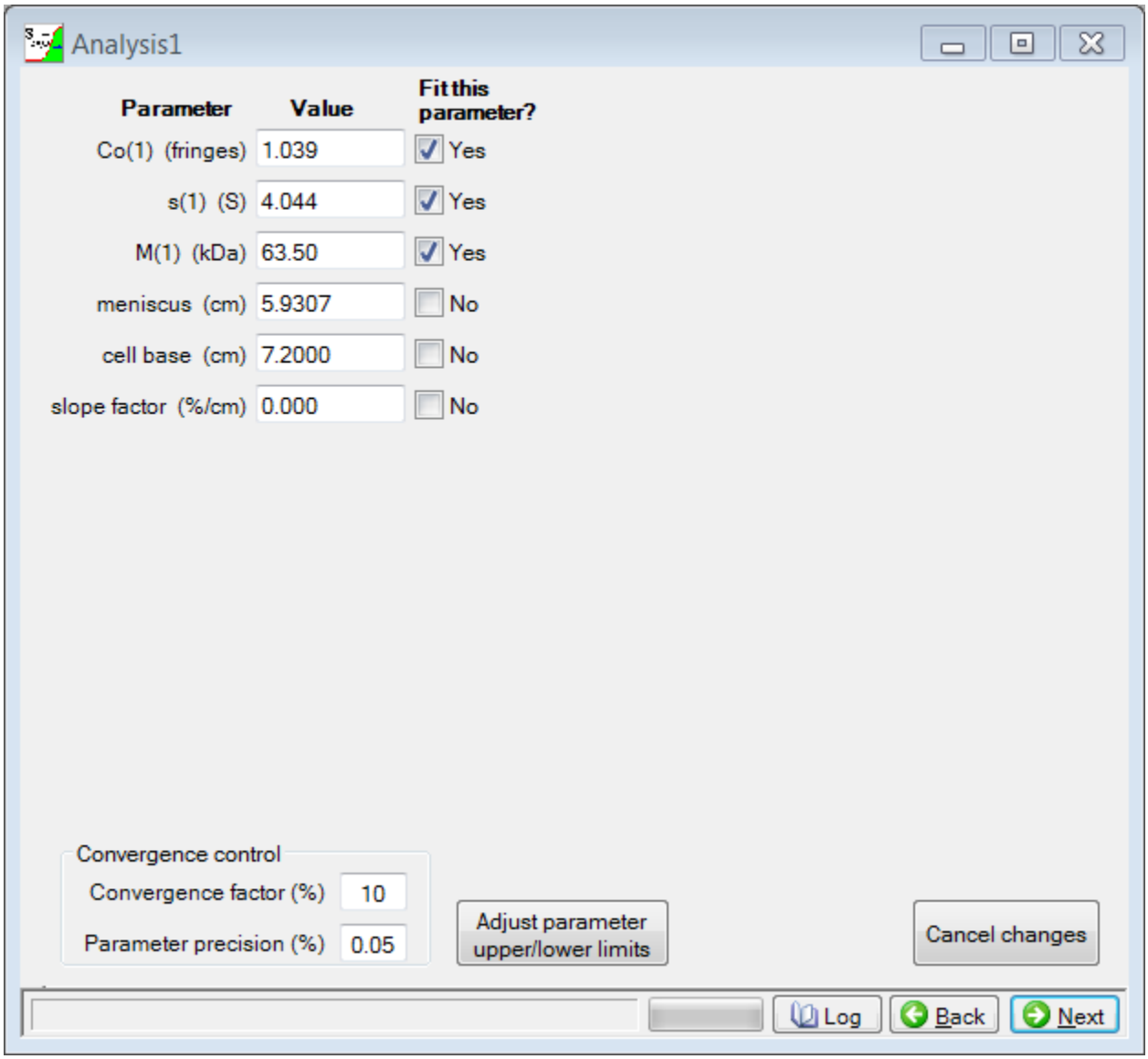
When new data have been loaded, upon the first entry to the **Set/alter fitting parameters page (Section 3.2.4)** you will see this dialog box:



(1) Answer 'Yes' and then

1. A program wizard will estimate the total loading concentration and enter that as the initial guess for the concentration fitting parameter.
    - If this were a multi-species fit the wizard assigns one tenth of the total to each of species 2 to N, and the balance to species 1.
2. Another wizard will estimate the overall average sedimentation coefficient for the sample and enter that as the initial guess for species 1.
    - If this were a multi-species fit the wizard assigns the sedimentation coefficient of species 1 as 0.9 times the overall average and then values for species 2 to N by assuming they are dimer, trimer, tetramer, hexamer, and octamer of species 1.
3. The program then enters an initial guess for the molar mass of each species based on its sedimentation coefficient and an assumed frictional coefficient ratio (f/f_0 ratio) of 1.4.

(2) Those initial guesses for concentration, sedimentation coefficient, and molar mass are then displayed as shown below so you can review and alter them, if needed, based on what you know about your sample. In this case we will simply use those initial guesses.



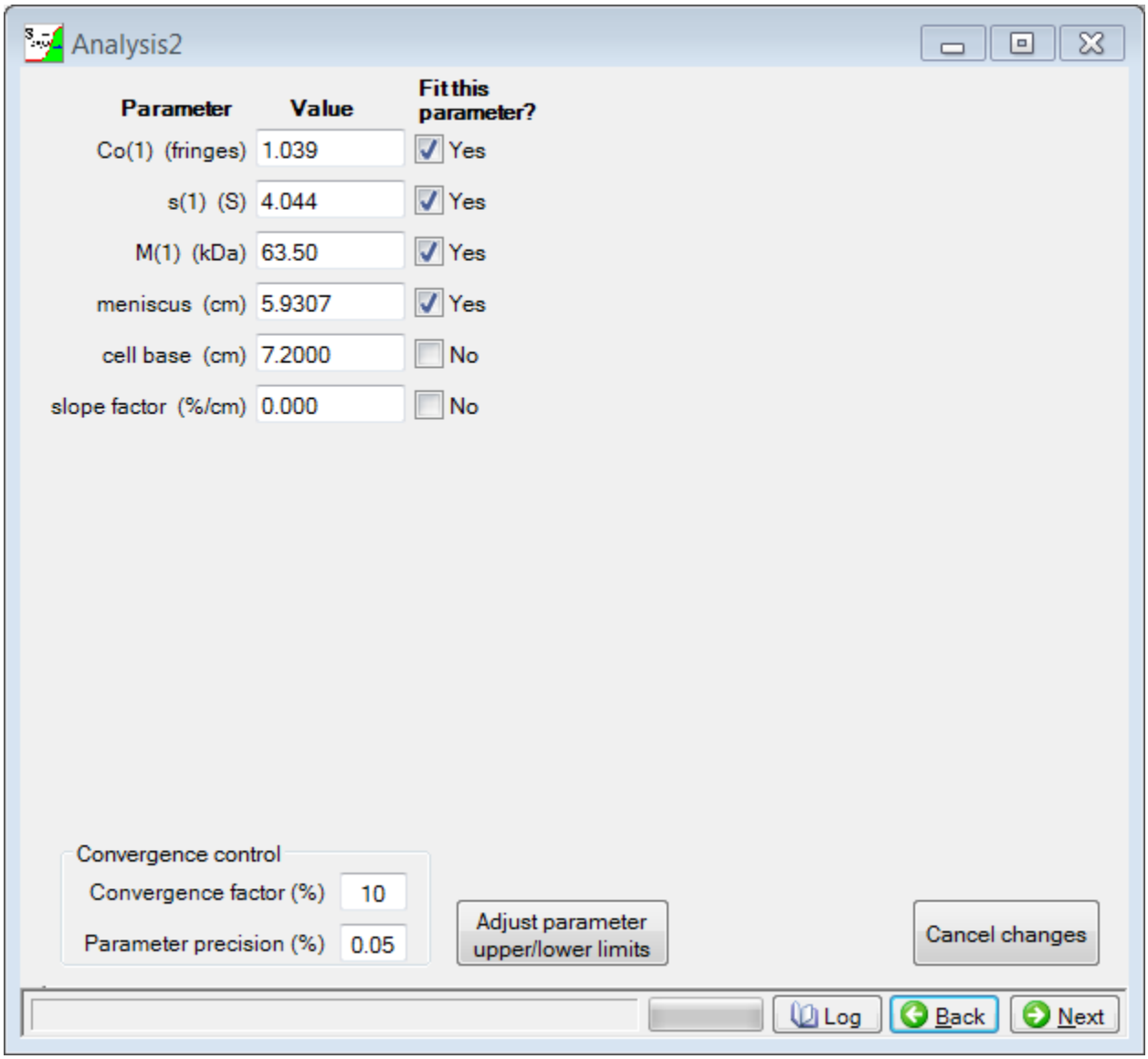
(3) This page also allows you to review (and possibly alter) which of the available fitting parameters will be fitted (versus being held fixed at the current value). Ones that you will likely want to confirm are whether to fit for the positions of the meniscus and the cell base.

The default is not to fit the meniscus position, so the check box to the right of that parameter is unchecked. However for interference data it is often difficult to judge the true meniscus position from the raw data (especially when you use meniscus-matching centerpieces) and therefore we will click the check box to turn on fitting of the meniscus position.

Similarly the fitting of the cell base position is not turned on by default. Since the portions of the cell we are fitting in this case appear to have only a small influence from accumulation at the cell base, initially we will hold the cell base fixed at the nominal position of 7.2 cm (but if the initial fit results suggest that may be a problem, we could try fitting it later).

(4) The controls in the lower left corner of this page relate to advanced options and seldom need to be adjusted. See **Convergence control (Section 3.2.4.4)** and **Adjust parameter upper/lower limits (Section 3.2.4.5)** for details.

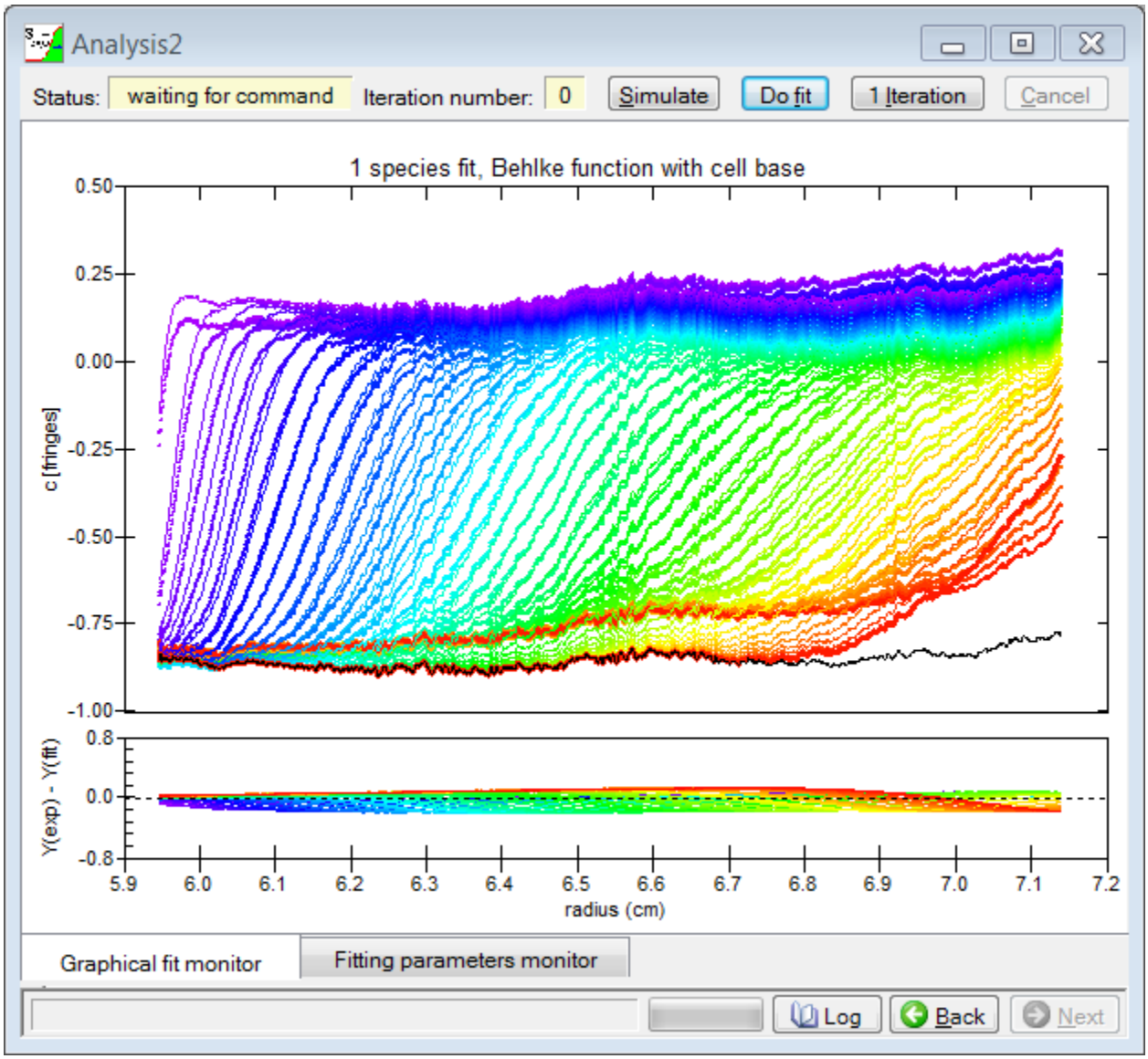
Our final setup looks like this:



We are ready to continue so push the **Next button (Section 3.1.3.1)** to move to the **Perform least-squares fit page (Section 3.2.5)**.

5. Perform the fit

Upon entering the **Perform least-squares fit page (Section 3.2.5)** the program calculates the fitting function using the initial guesses we just established and displays this graph with an overlay of that result with the data being fitted, with a residual plot below.

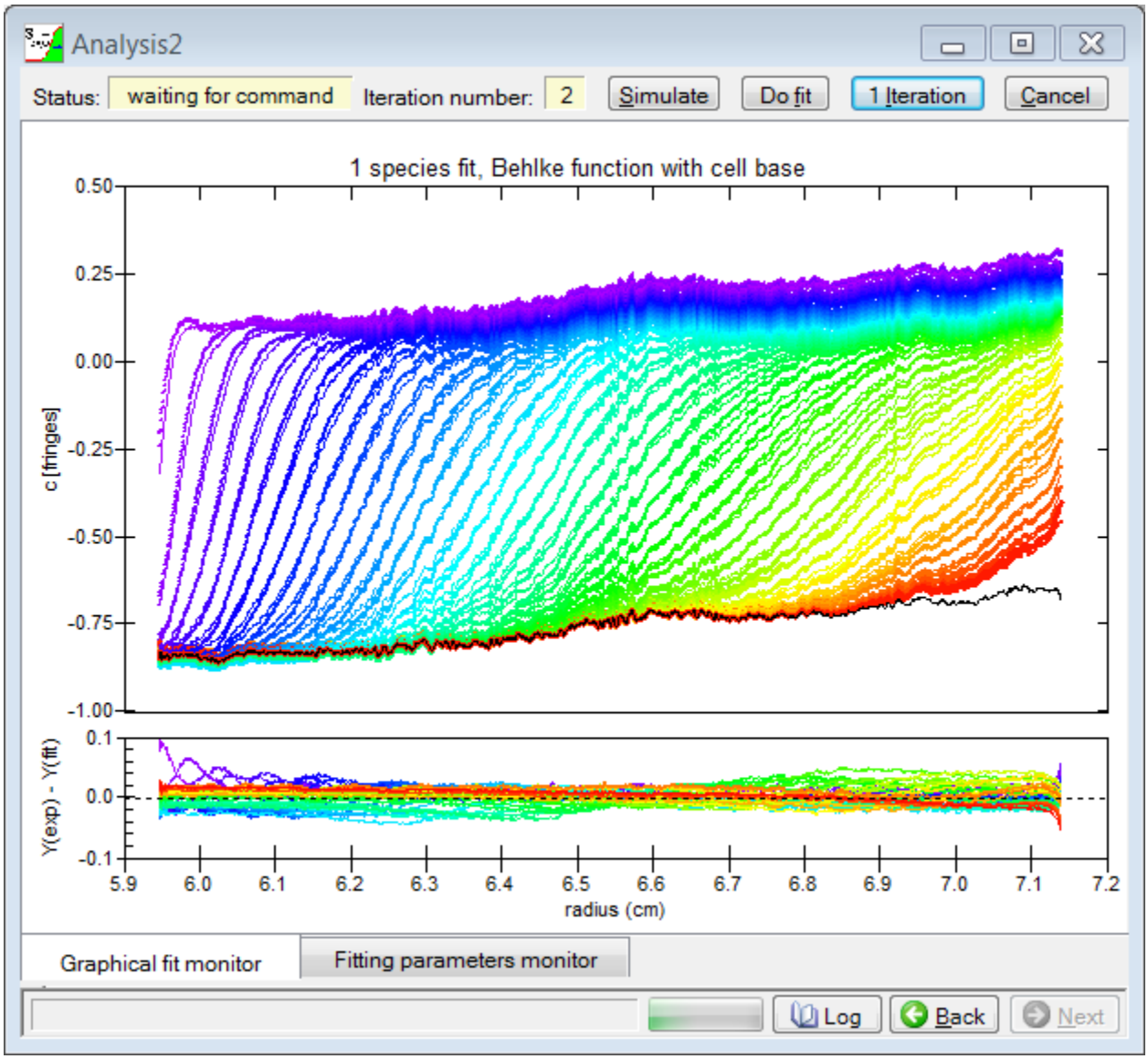


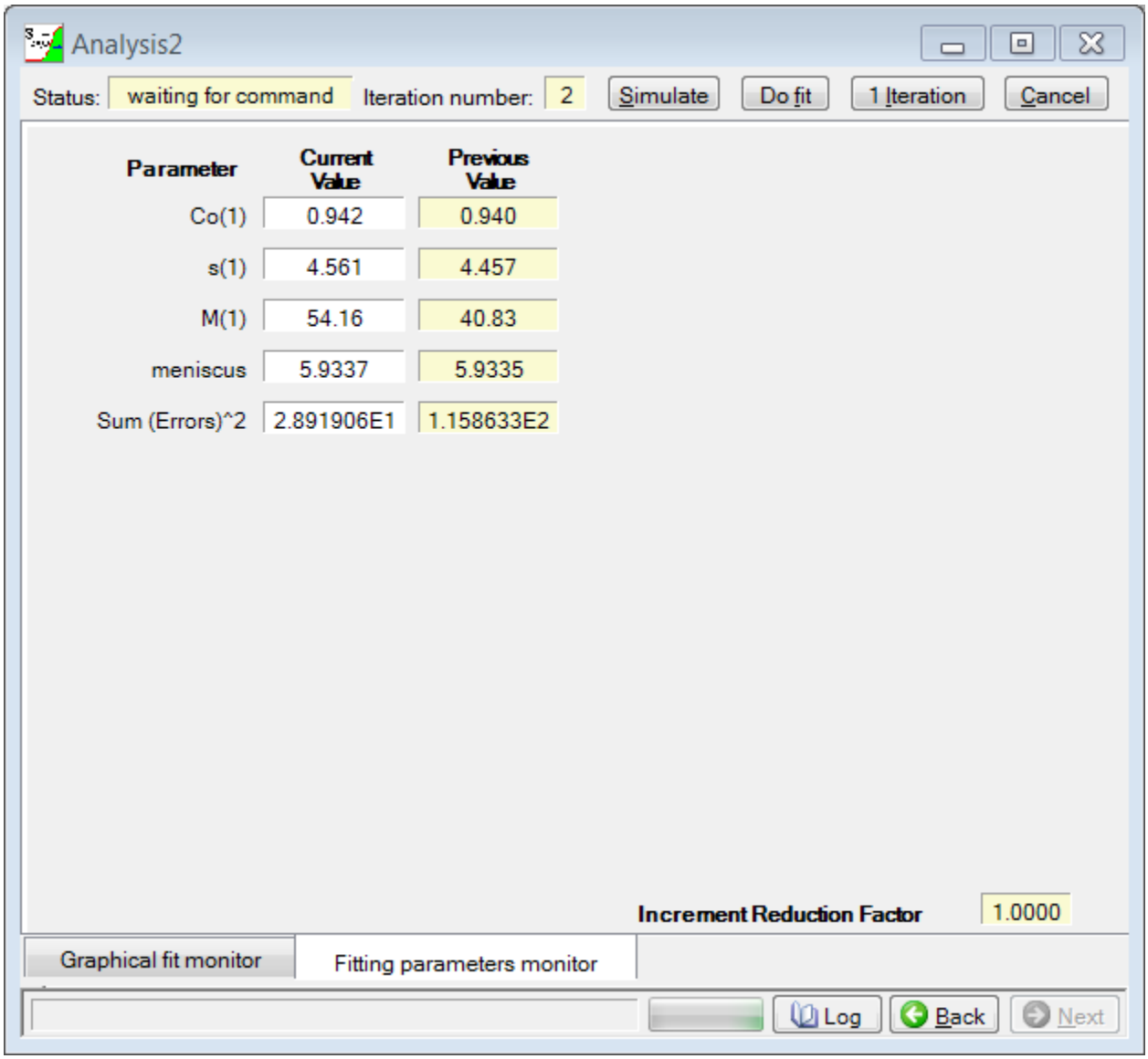
Clearly the initial guesses by the wizard do not provide a good fit, but they are quite good enough for our purpose.

Usually at this point you will simply push the **Do fit** button (which is selected by default so you can just hit **Enter** from the keyboard instead). As the fit proceeds the graph will be updated and you will see the fitted curves move closer to the experimental ones and the residuals will diminish (the vertical scaling of the residual plot may change as the peak-to-peak excursion falls). If for some reason you want to halt the fit before it finishes use the **Cancel** button (or the **Back button (Section 3.1.3.1)**)

If you wish to monitor the fit progression more closely you can optionally step through the fitting procedure one iteration at a time using the **1 iteration** button. You also can choose to follow the actual changes in parameter values rather than the graph by selecting the **Fitting parameters monitor tab (Section 3.2.5.6)**.

The following 2 images show the **Graphical fit monitor** and **Fitting parameters monitor** tabs after 2 iterations (2 pushes of the **1 iteration** button):


Analysis2
Status: waiting for command
Iteration number: 2
Simulate
Do fit
1 Iteration
Cancel
1 species fit, Behlke function with cell base
c [fringes]
0.50
0.25
0.00
-0.25
-0.50
-0.75
-1.00
Y(exp) - Y(fit)
0.1
0.0
-0.1
5.9
6.0
6.1
6.2
6.3
6.4
6.5
6.6
6.7
6.8
6.9
7.0
7.1
7.2
radius (cm)
Graphical fit monitor
Fitting parameters monitor
Log
Back
Next



The graphs show substantial improvement in the overlap of the fitted curves to the raw data and a substantial reduction in the magnitude of the residuals.

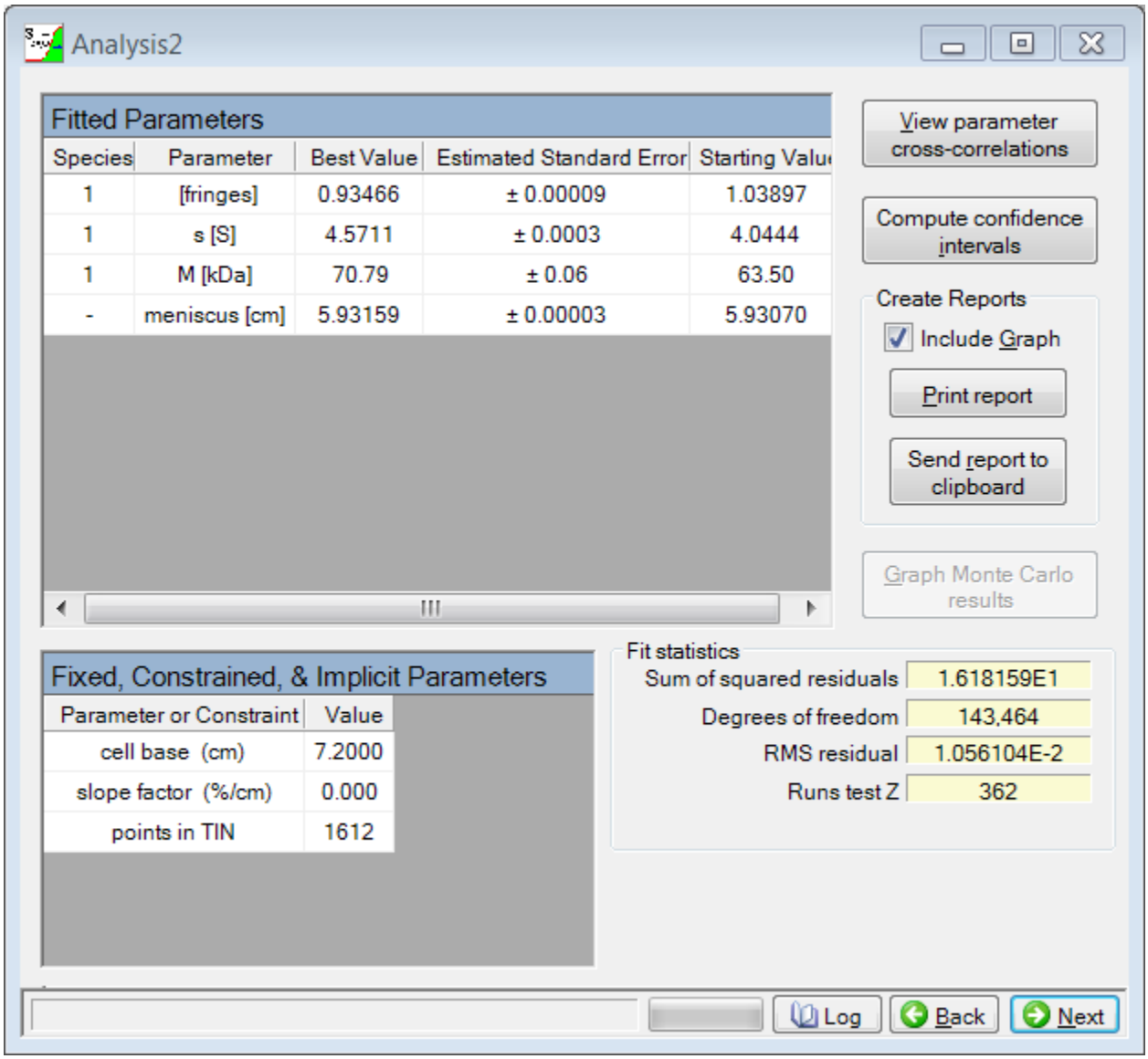
Note that the **Fitting parameters monitor** tab shows only those parameters actually being fitted (not the ones being held constant, such as the cell base position in this case). From this display we can see that the only parameter which changed substantially between the first and second iterations was the molar mass, and that the second iteration reduced the sum-of-squared residuals (variance) nearly 4-fold.

Also note that if you are using the **1 iteration** button to halt after each iteration, at any point you can push the **Do fit** button to complete the fit without further pauses.

When the fit completes (converges) the program will automatically advance to the **Report fit results page (Section 3.2.6)**.

6. Review the fit results

When this fit completes (which takes 6 total iterations) the **Report fit results page (Section 3.2.6)** looks like this:



The **Fitted parameters table** at the upper left lists the best-fit parameters, the starting value for that parameter (the initial guess), and a quick estimate of the statistical standard error for that parameter (its precision).

The **Fixed, constrained & implicit parameters table** at the lower left lists the values of any fixed (non-fitted) parameters. Since we are fitting the TIN the number of data points (fitting parameters) used to fit the TIN is noted. For multi-species fits this table may also show constraints that have been imposed to link properties of different species (see **Constrain M or s ratios among species (Section 3.2.3.7)**) or values which can be implicitly calculated from the fitting parameters (for example the total loading concentration).

The **Fit statistics** group at the lower right lists a statistical summary of the fit, including the sum of the squared *residuals*, the *degrees of freedom*, the *rms residual*, and the *runs test Z* score. Note that although a rms residual of 0.0106 fringes is not as low as the intrinsic noise in a well-adjusted instrument (0.005 fringes or less), real experiments often don't approach that level, so this value is not necessarily an indication that this is a bad fit.

This is a good time to point out the **Log page (Section 3.1.3.2)**, which we can examine at any time using the **Log button** along the bottom of the **Analysis window (Section 3.1.3)**. It has automatically recorded everything we have done so far, including the results and a summary graph of this fit. Here is the contents of that log (which can be copied and pasted into your electronic notebook or word processor):

[10/6/2011; 13:54:24]

New analysis started using program version 7.0.0.

[10/6/2011; 13:54:38]

Loaded 90 new scans from folder C:\XLAWIN\PureBSA_IP_sim\with noise:

00001.IP1, 00002.IP1, 00003.IP1, 00004.IP1, 00005.IP1, 00006.IP1, 00007.IP1, 00008.IP1, 00009.IP1, 00010.IP1, 00011.IP1, 00012.IP1, 00013.IP1, 00014.IP1, 00015.IP1, 00016.IP1, 00017.IP1, 00018.IP1, 00019.IP1, 00020.IP1, 00021.IP1, 00022.IP1, 00023.IP1, 00024.IP1, 00025.IP1, 00026.IP1, 00027.IP1, 00028.IP1, 00029.IP1, 00030.IP1, 00031.IP1, 00032.IP1, 00033.IP1, 00034.IP1, 00035.IP1, 00036.IP1, 00037.IP1, 00038.IP1, 00039.IP1, 00040.IP1, 00041.IP1, 00042.IP1, 00043.IP1, 00044.IP1, 00045.IP1, 00046.IP1, 00047.IP1, 00048.IP1, 00049.IP1, 00050.IP1, 00051.IP1, 00052.IP1, 00053.IP1, 00054.IP1, 00055.IP1, 00056.IP1, 00057.IP1, 00058.IP1, 00059.IP1, 00060.IP1, 00061.IP1, 00062.IP1, 00063.IP1, 00064.IP1, 00065.IP1, 00066.IP1, 00067.IP1, 00068.IP1, 00069.IP1, 00070.IP1, 00071.IP1, 00072.IP1, 00073.IP1, 00074.IP1, 00075.IP1, 00076.IP1, 00077.IP1, 00078.IP1, 00079.IP1, 00080.IP1, 00081.IP1, 00082.IP1, 00083.IP1, 00084.IP1, 00085.IP1, 00086.IP1, 00087.IP1, 00088.IP1, 00089.IP1, 00090.IP1

[10/6/2011; 13:54:40]

Meniscus position set to 5.9307 cm.

[10/6/2011; 13:54:40]

Automatic jitter correction performed in the region from 5.88 to 5.8941 cm.

[10/6/2011; 13:54:40]

Automatic fringe shift removal performed using the 'Middle half' option.

[10/6/2011; 13:54:41]

Current fitting model is 1 species, using the Behlke function with buildup at the cell base. The species molecular parameters are sedimentation coefficient and molecular mass (in true mass units). Baseline noise will be removed by fitting the TIN.

[10/6/2011; 13:54:50]

Made new initial guesses for parameters based on an estimated sedimentation coefficient of 4.04 S and a loading concentration of 1.04 fringes.

[10/6/2011; 14:04:16]

Fitting of parameter meniscus was turned on.

[10/6/2011; 14:07:35]

Single fit iteration started.

[10/6/2011; 14:08:07]

Fit iteration # 1 manually started.

[10/6/2011; 14:52:30]

Full fit started...

[10/6/2011; 14:52:40]

Fit Completed (6 iterations).

Results of fit:

Parameter	Best Value	Standard Error	Starting Value
Co(1) (fringes)	0.93466	+/- 0.00009	1.03897
s(1) (S)	4.5711	+/- 0.0003	4.0444
M(1) (kDa)	70.79	+/- 0.06	63.50
meniscus (cm)	5.93159	+/- 0.00003	5.93070

Parameters held constant:

Parameter	Fixed Value
cell base (cm)	7.2000
slope factor (%/cm)	0.000

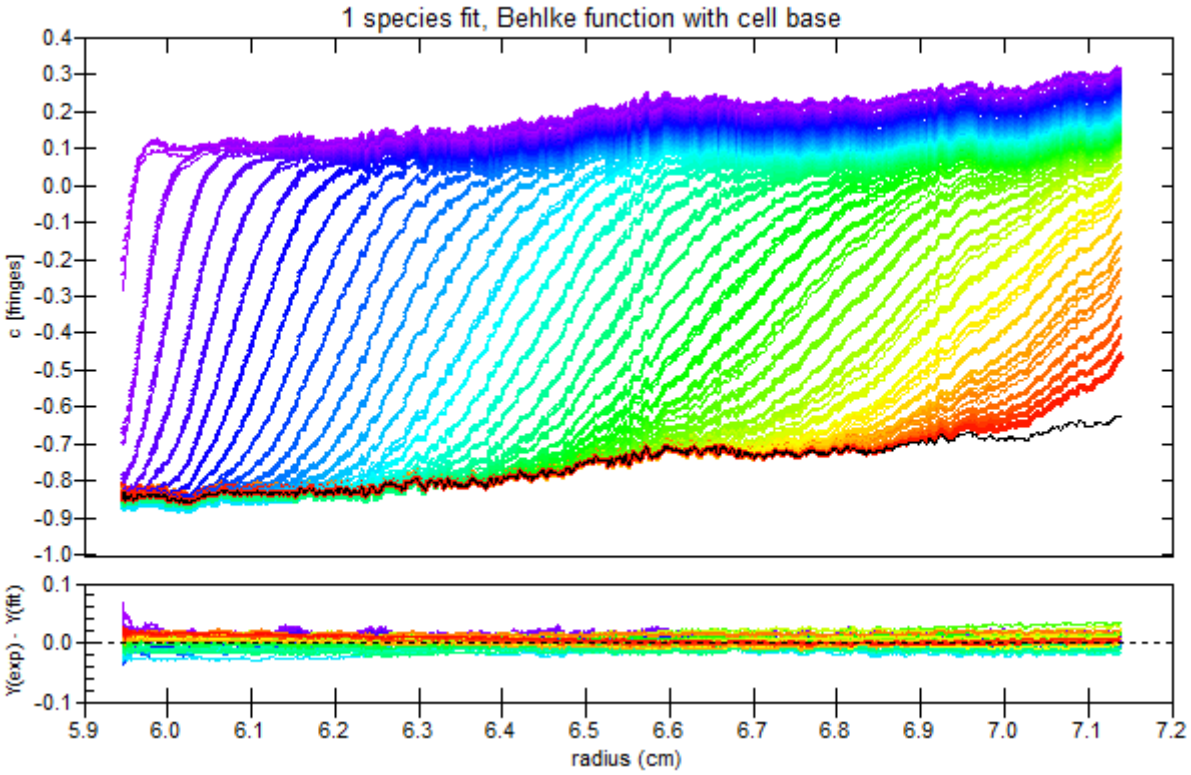
Fit statistics and convergence settings:

Final sum of squared residuals = 16.18159 (starting value = 878.8474); rms = 0.01056104, for 145,080 data points and 143,464 degrees of freedom; runs test Z = 362.

Convergence limit = 0.1; parameter precision limit = 0.05%.

Sedimentation coefficients are uncorrected (raw) values.

The diffusion coefficient corresponding to the best fit molecular mass is 5.83 F.

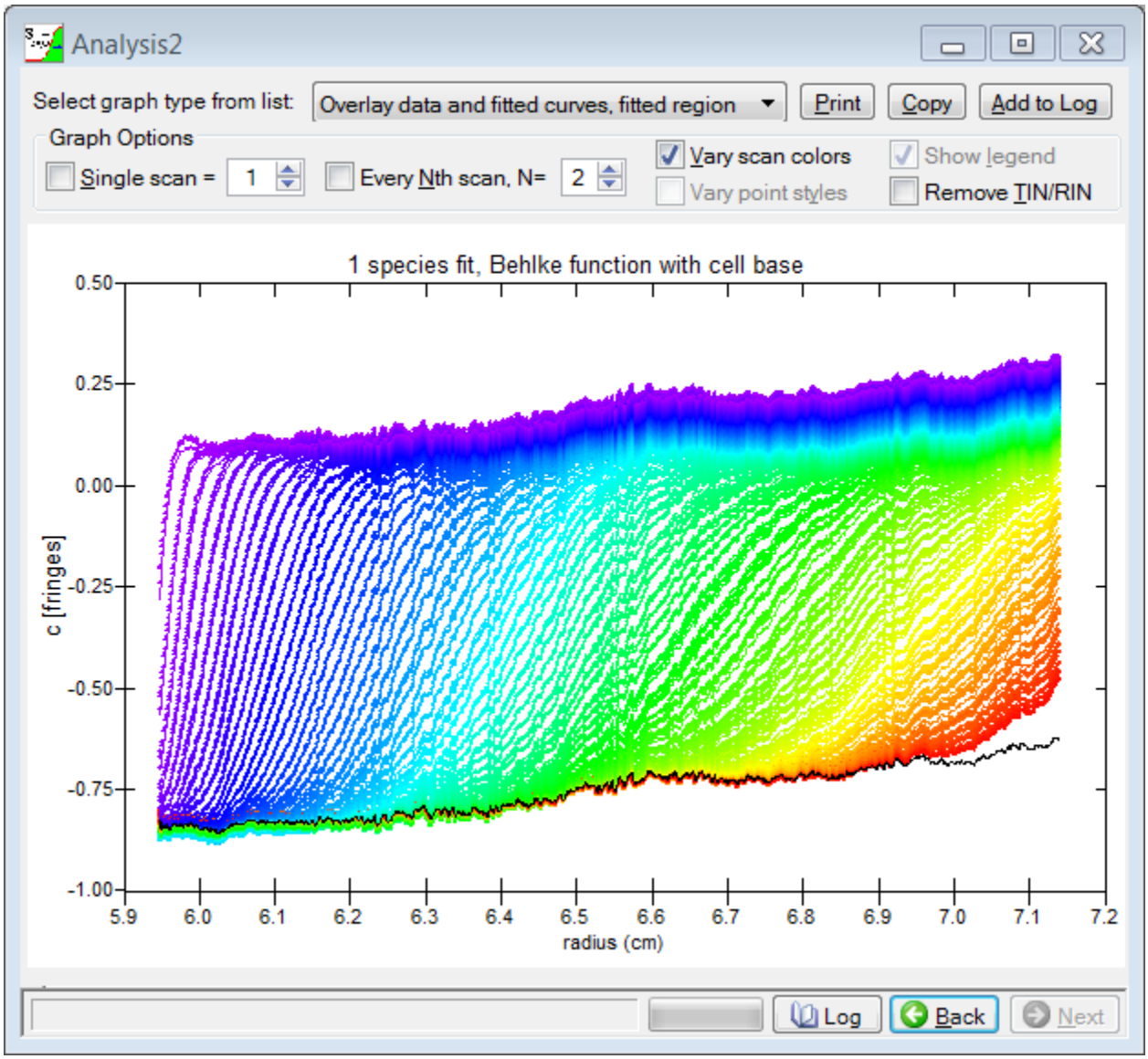


Clearly the overlay and residual plots in the Log above suggest this is a reasonably good fit. But before doing anything else it is probably a good idea to save what we've done so far to disk in case we want to return to this point, which can be done using either the **Save button** on the **Toolbar (Section 3.1.2)** or the **File...Save current analysis as...** menu item.

Often after completing a fit you will want to examine various graphs of the fitting results. Push the **Next button (Section 3.1.3.1)** to move to the **Graph fit results page (Section 3.2.7)** (the last page).

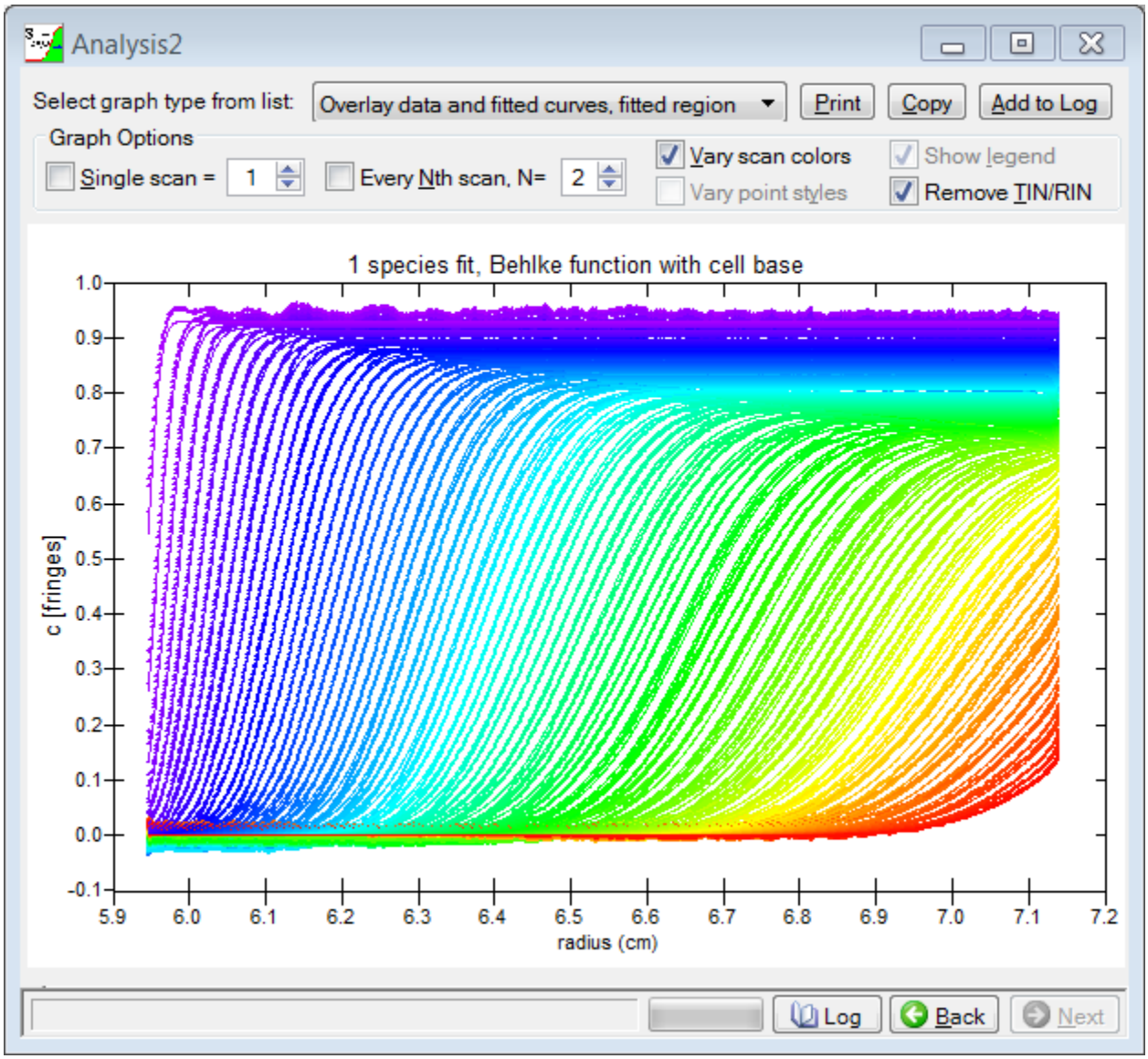
- **7. Examine graphs of the fitting results---is this a good fit?**

After the first fit in each analysis window this will automatically display an overlay of the fitted data and the best-fit curves, as shown below:



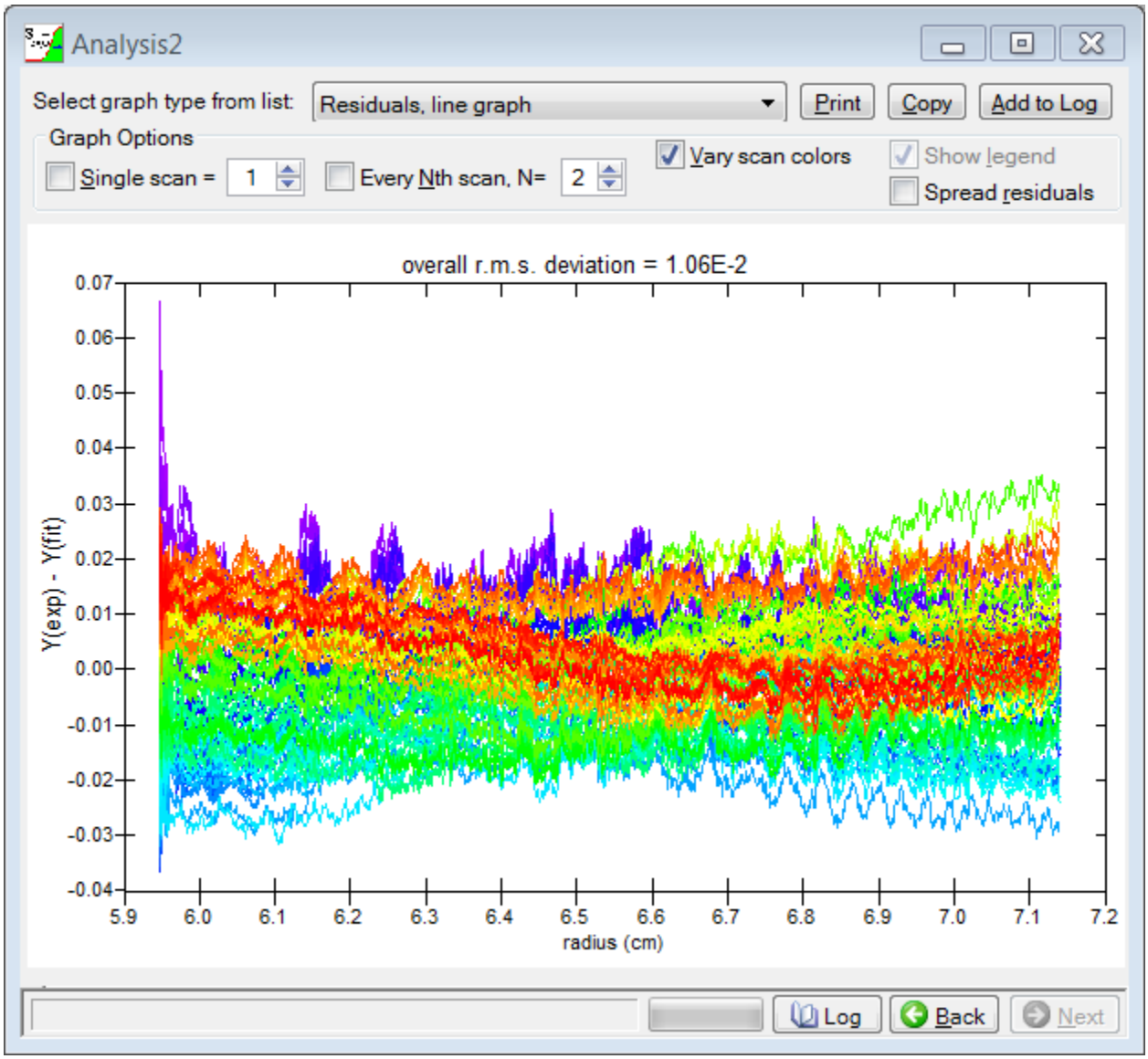
Obviously what we would like to see is a perfect overlay of fitted curves with the raw data (except for deviations due to random noise). Clearly in this case we do not have a perfect overlay, and some of the deviations appear to be systematic. However it is important to remember that for interference scans the noise is definitely not random and may be systematically positive or negative over certain regions of the cell for certain groups of scans.

Note that the black line near the bottom of the graph is the TIN (the fitted baseline profile). If we want to see the data and fit after removing this systematic noise (to give cleaner boundaries) that can be done by checking the **Remove TIN/RIN** checkbox at the right above the graph, giving this result:



Various other options available for this type of graph (and all the different graphs) are detailed **here (Section 3.2.7)**.

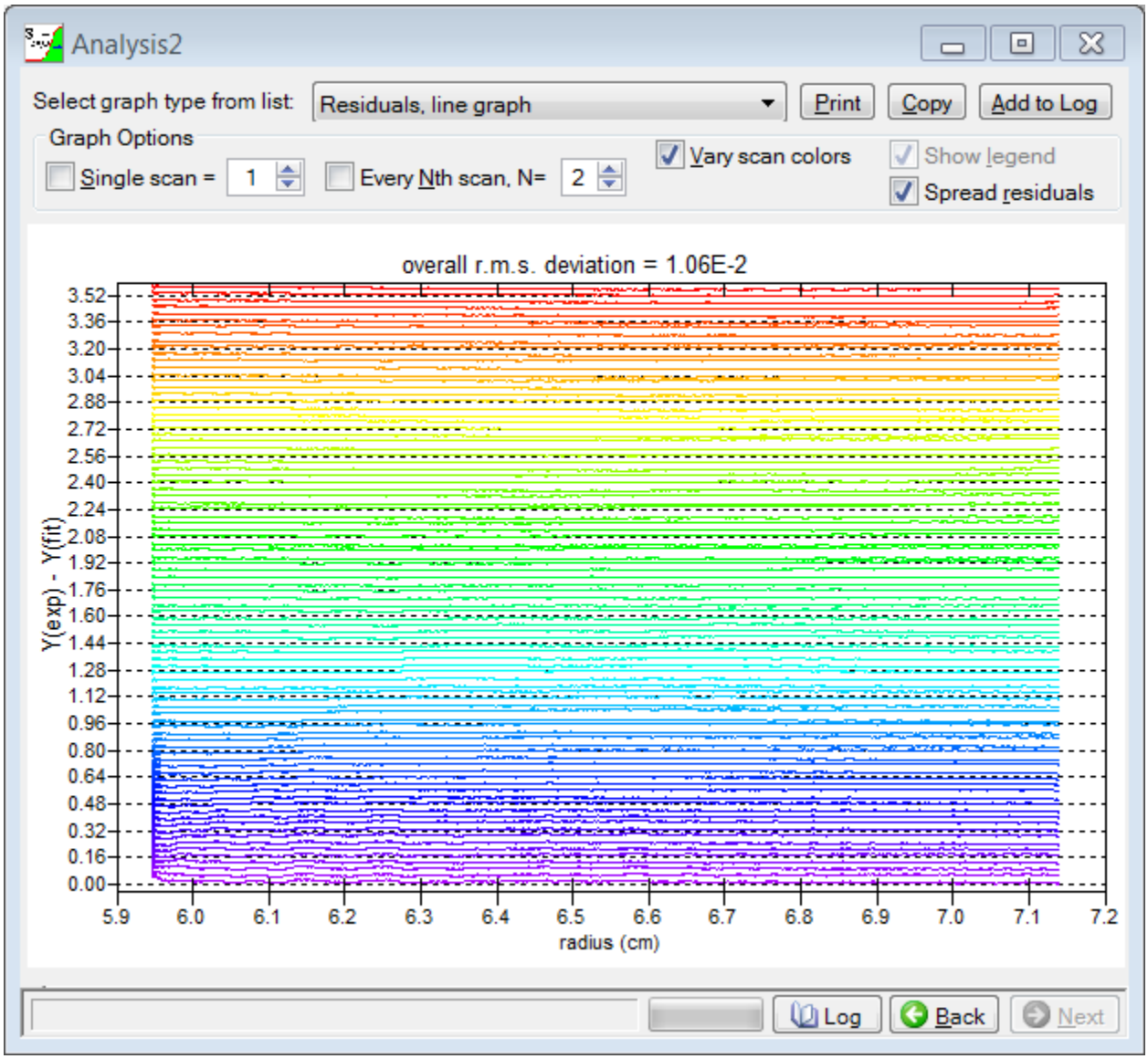
Residual plots generally provide a more sensitive test of the quality of the fit than the overlay plots. Two fundamentally different types of residual plots are available. One is a standard line plot, as shown below (the different plot types are selected using the drop-down list box at the top of the page).



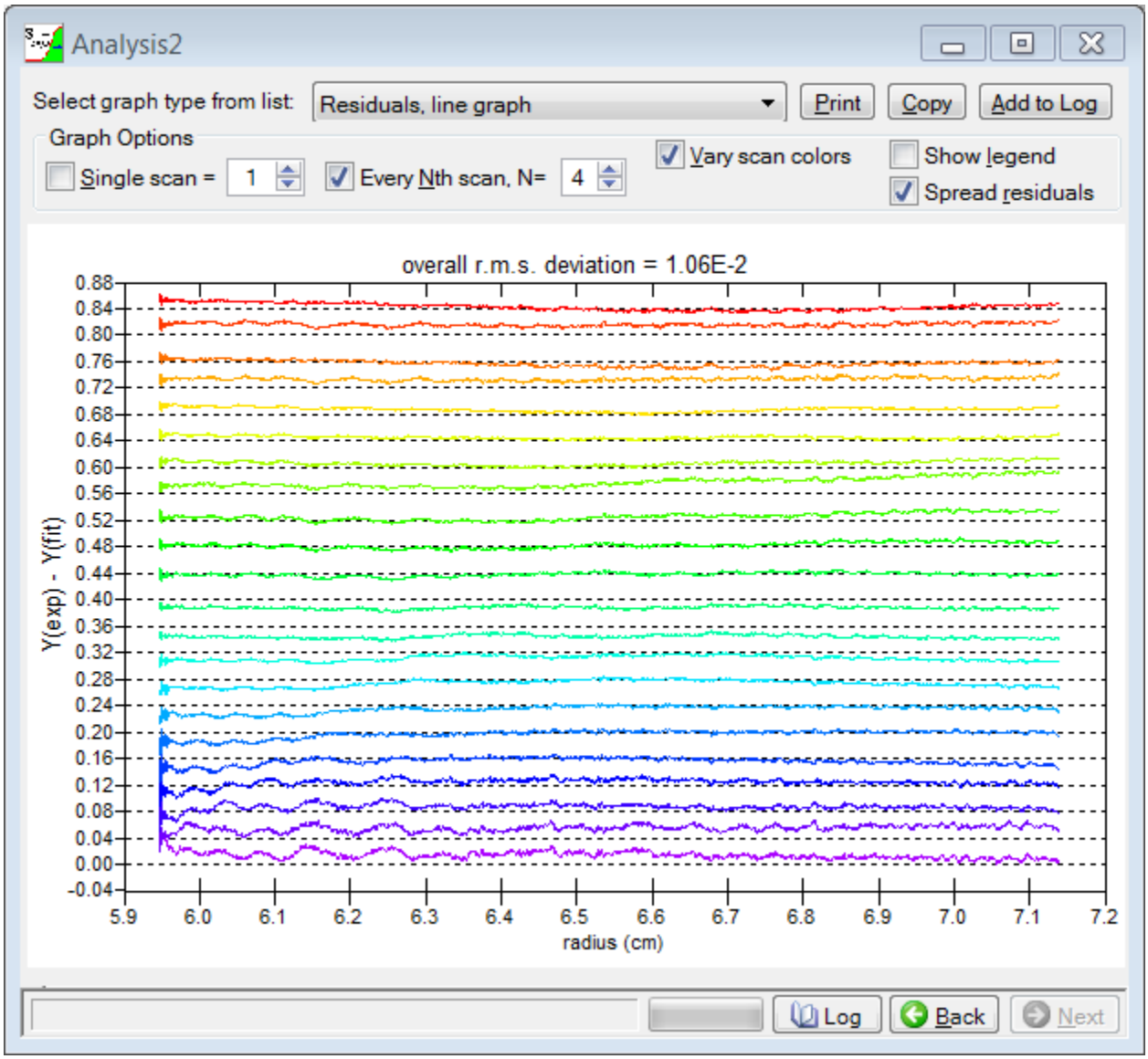
Ideally of course what we would like to see is simply noise randomly distributed around zero.

Unfortunately when there are many scans, and especially when the fit is good and the noise is fairly random (as it usually is for absorbance scans) there are often so many scans overlaying each other it is hard to see what is going on. In this case the rainbow color coding helps us to see that in the left portion of the cell (nearer the meniscus) the deviations tend to be negative early in the run (bluer colors) and positive later in the run (redder colors). This could be an indication of a problem with the fit model (perhaps there is a minor species with a very low sedimentation coefficient), or perhaps with the experiment itself (there was not a good match between the sample and the reference buffer so we are detecting sedimentation of buffer components).

An alternative form of this plot tries to overcome the data overlap problem by shifting each successive scan upward by a constant amount (the **Spread residuals** option shown at the right above the graph). In this case this gives the following graph:



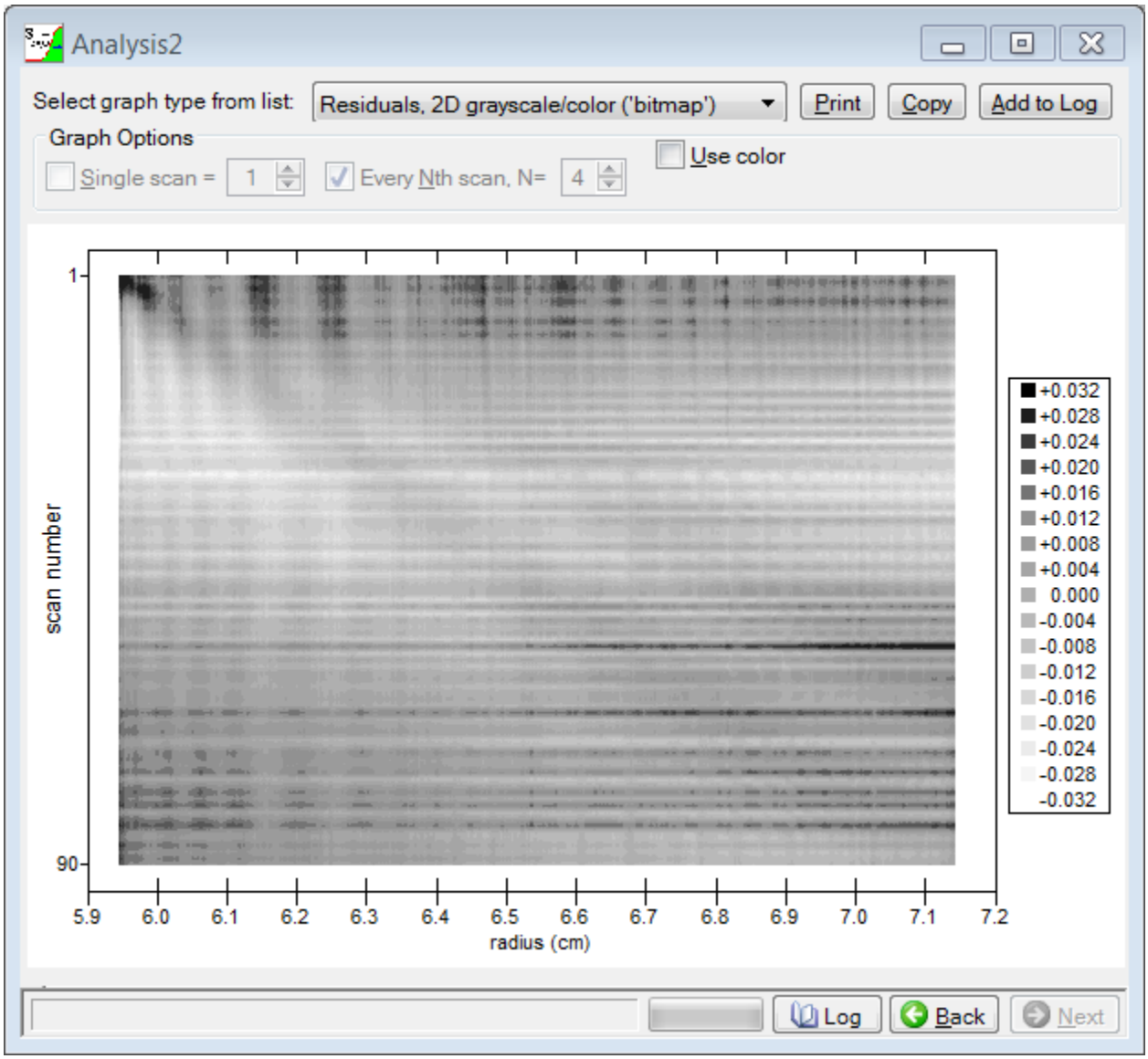
However with 90 scans this option does not work too well. A better choice here would be to show only every fourth scan, which we can do by checking the **Every Nth scan** check box and setting N to 4 using the adjacent up/down spinner. This gives the following graph:



With this reduced number of scans, the vertical offset between successive traces can correspond to the spacing of the tic marks along the Y axis, and thus ideally each residual plot will lie exactly along the dotted lines. It is now possible to see that some of the traces are systematically shifted above or below those dotted lines (the residuals are all above or below zero for that scan). That suggests that we may be having a problem achieving complete removal of the *jitter* or RIN component of systematic noise.

We can also now see that the earlier scans (lower and bluer) for some reason show a higher amplitude of the wave-like 'ripples' which are a characteristic form of systematic noise in interference scans.

The second fundamental type of residual plot is a 2-dimensional gray scale contour plot, similar to the 'bitmap' residual plot in SEDFIT, with radius along the X axis and scan numbers along the Y axis (with the first scan at the top to follow the convention used in SEDFIT). The magnitude of each residual determines the shade of gray (or optionally color) shown at that position. Here is that plot for this fit:



In this representation, a bad fit will produce dark and light bands which move diagonally downward and to the right (along with the boundaries in the cell). A weak such pattern is visible above.

The horizontal dark and light bands visible here indicate that the noise in these interference scans is changing its characteristics from one scan (or group of scans) to another.

The presence of vertical (or near-vertical) dark and light bands indicates that the fitted TIN is not completely removing the systematic baseline profile (in other words the baseline is not actually completely constant in time). The slight bending of these bands over the middle portion of the run might be due to small radial shifts of the rotor position relative to the optics.

A comparison of the above graph with the preceding two illustrates that our brains are much better at detecting subtle patterns in the gray scale representation versus the line graphs.

Note that with absorbance scans these plots look quite different (because the noise is much more random). Go **here (Section 3.2.7)** for an example.

8. Refining the fit

Clearly an important question is whether this is a good fit. A number of aspects about this question are discussed in more depth in the FAQ **What constitutes a good fit? (Section 5.1.7)** Another aspect of this question is to ask "what more could we do within this single-species model, or what could we do differently

that might improve the fit"?

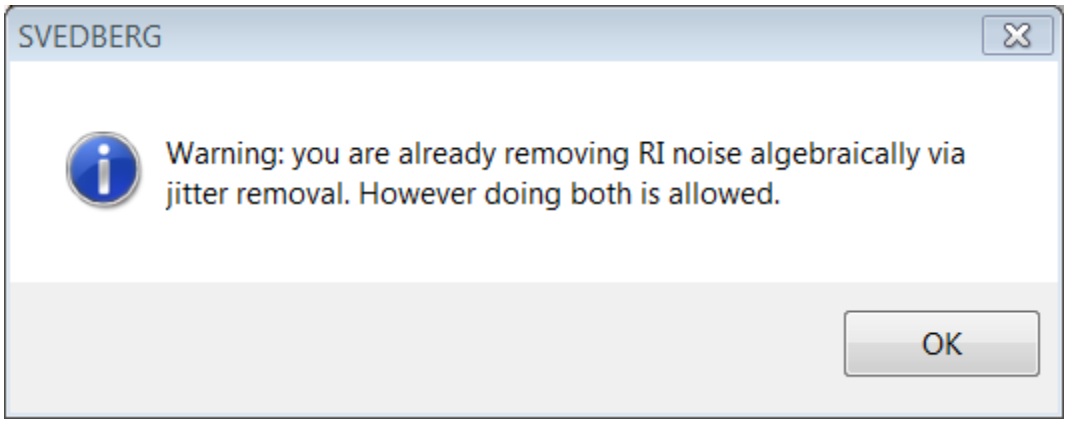
The only relevant fitting parameter we have not already optimized is the position of the base of the cell. Certainly we could try fitting that parameter, but as noted when we set up this fit it does not appear that for this experiment the cell base position has much influence on the data in the region being fitted.

However we did see from the residual plots in the preceding section some evidence that the *jitter* or RIN might not have been optimally removed. In the current fit the jitter was handled by aligning the fringes in the air-air space, so we should perhaps try the alternative approach of including the RIN (a vertical offset for each scan) as a fitting parameter.

To do that we need to navigate back to the **Select fitting model page (Section 3.2.3)**, which can be done 3 different ways:

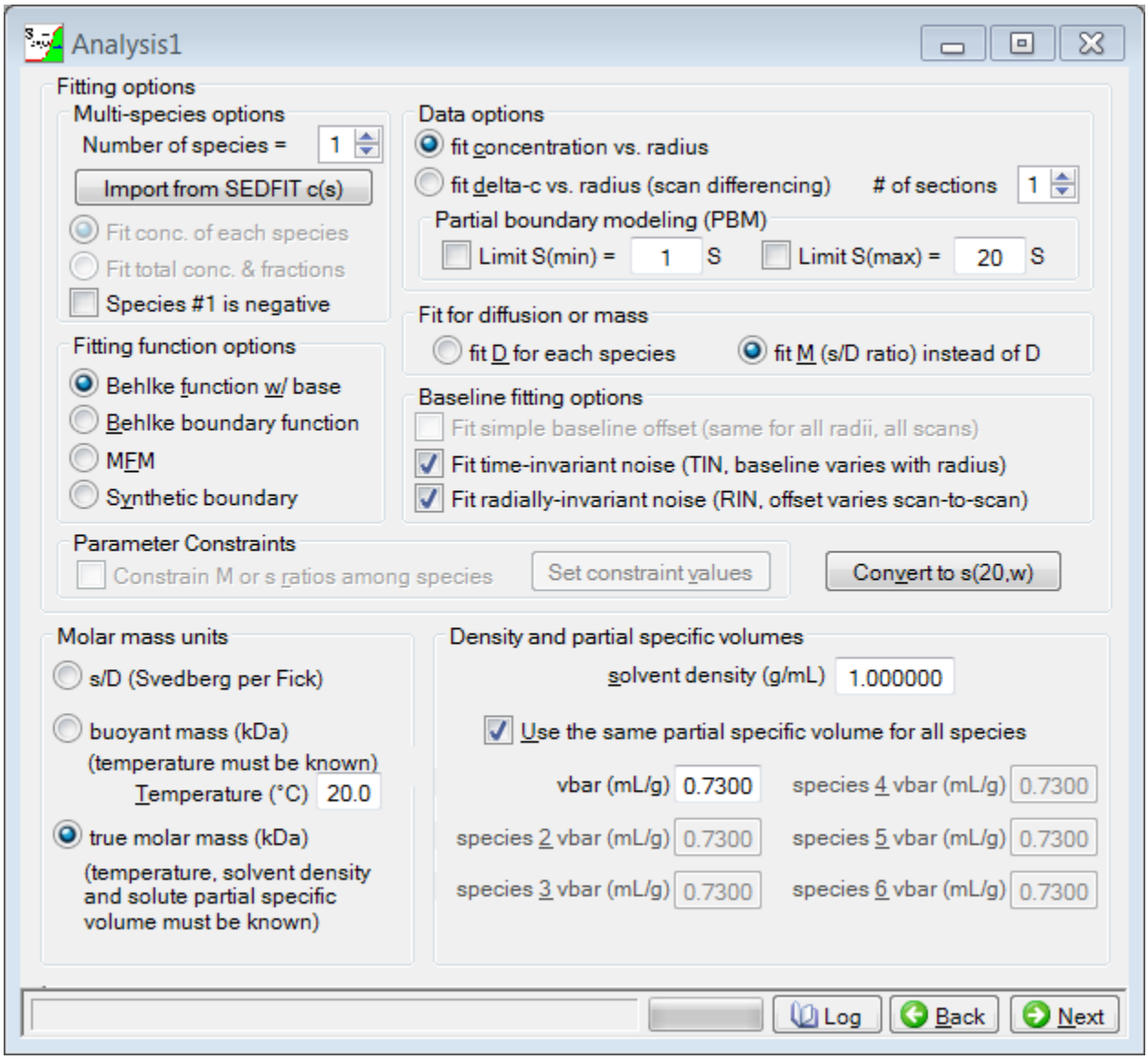
1. pushing the **Back button (Section 3.1.3.1)** 4 times (assuming you are still on the **Graph fit results page (Section 3.2.7)**)
2. clicking on the 'Select fitting model' node within the **Navigation tree (Section 3.1.4)**
3. right-clicking on any blank portion of the **Analysis window (Section 3.1.3)** (not on the graph or a control) to bring up a context menu and selecting the appropriate page from the 'Go to page' submenu

Once reaching the **Select fitting model page (Section 3.2.3)**, to turn on fitting of the RIN simply click the appropriate check box within the **Baseline fitting options** group box. When you do so you will see this reminder message:

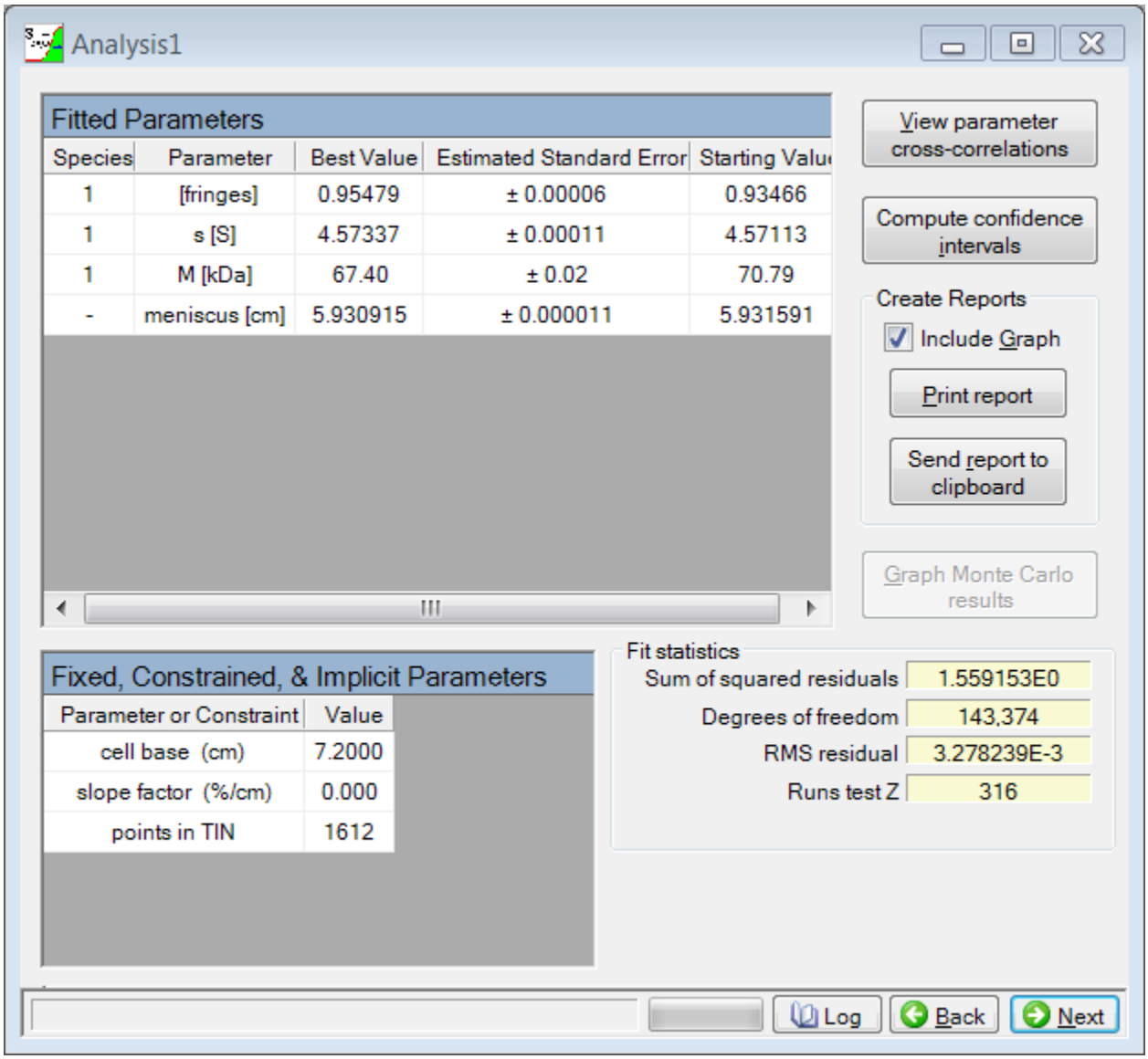


Using both RIN removal methods together is allowed; the only thing that changes are the actual values of the fitted RIN.

The fitting model page now looks like this:


Analysis1
Fitting options
Multi-species options
Number of species = 1
Import from SEDFIT c(s)
Fit conc. of each species
Fit total conc. & fractions
Species #1 is negative
Fitting function options
Behlke function w/ base
Behlke boundary function
MEM
Synthetic boundary
Data options
fit concentration vs. radius
fit delta-c vs. radius (scan differencing)
# of sections 1
Partial boundary modeling (PBM)
Limit S(min) = 1 S
Limit S(max) = 20 S
Fit for diffusion or mass
fit D for each species
fit M (s/D ratio) instead of D
Baseline fitting options
Fit simple baseline offset (same for all radii, all scans)
Fit time-invariant noise (TIN, baseline varies with radius)
Fit radially-invariant noise (RIN, offset varies scan-to-scan)
Parameter Constraints
Constrain M or s ratios among species
Set constraint values
Convert to s(20,w)
Molar mass units
s/D (Svedberg per Fick)
buoyant mass (kDa)
(temperature must be known)
Temperature (°C) 20.0
true molar mass (kDa)
(temperature, solvent density and solute partial specific volume must be known)
Density and partial specific volumes
solvent density (g/mL) 1.000000
Use the same partial specific volume for all species
vbar (mL/g) 0.7300
species 4 vbar (mL/g) 0.7300
species 2 vbar (mL/g) 0.7300
species 5 vbar (mL/g) 0.7300
species 3 vbar (mL/g) 0.7300
species 6 vbar (mL/g) 0.7300
Log
Back
Next

Moving forward and re-fitting the data with this slightly modified model gives the following result:

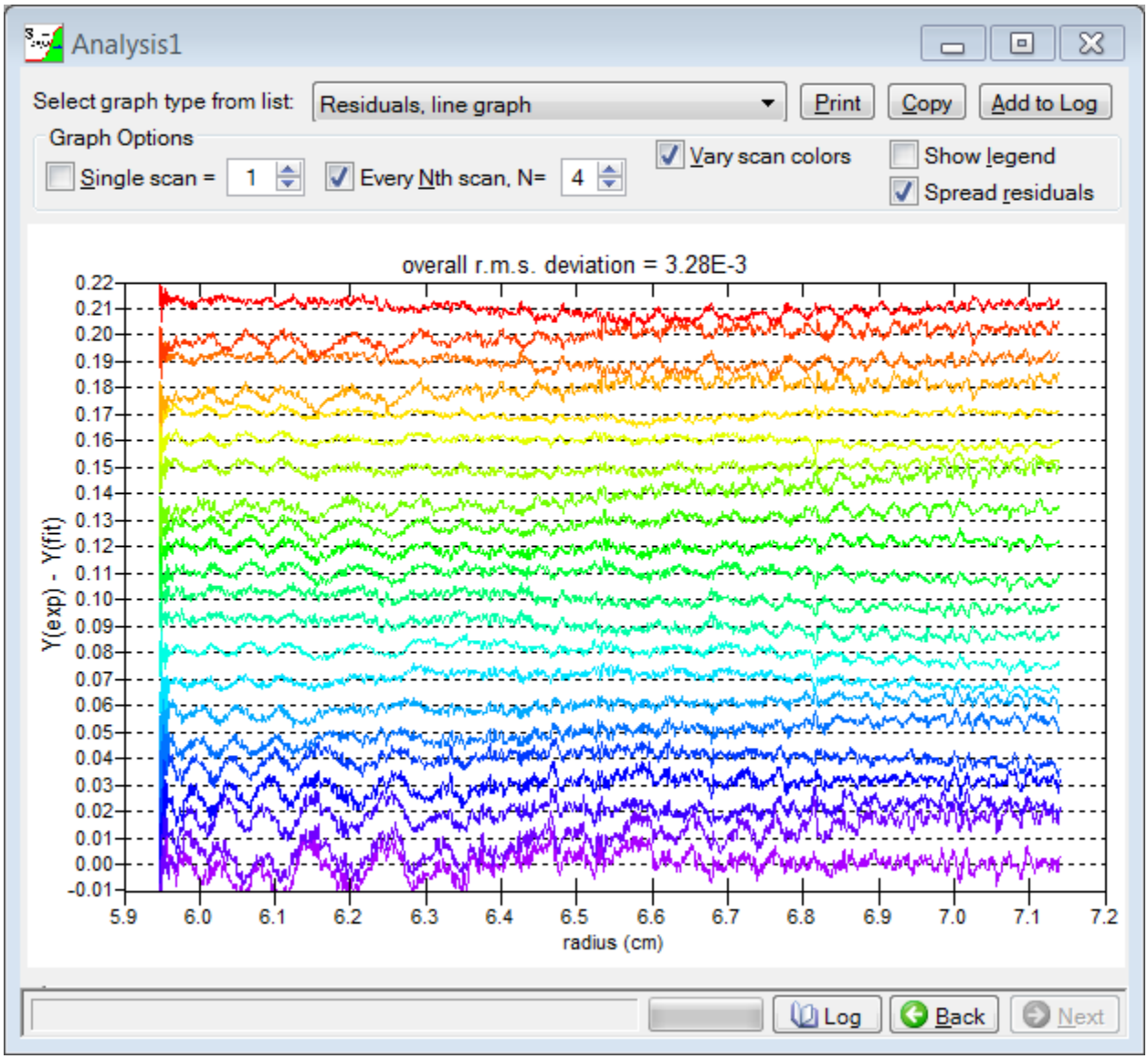


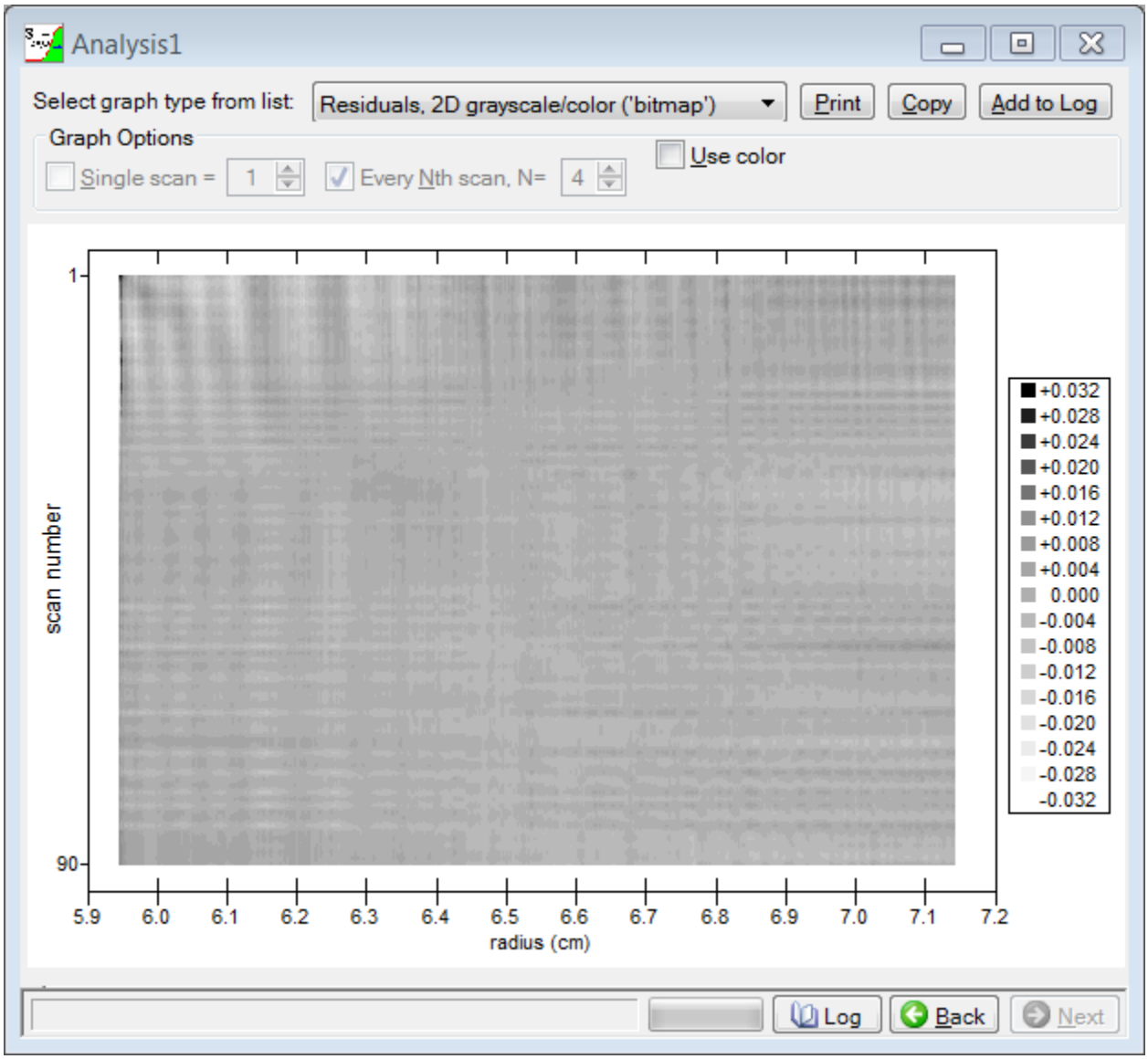
Comparing this to the earlier result (section 6 above) shows that fitting the RIN has reduced the rms residual more than 3-fold to a good value of 0.0033 fringes. What this means is that in reality the noise is not shifting the fringes uniformly up and down. The 'jitter' also often causes the fringes to tilt, and when that is true extrapolating the RIN based on the signal in the air-air space does not work well.

Note too that fitting the RIN has caused the best-fit loading concentration to increase about 2%. One drawback of fitting the RIN is that it makes the zero level for concentration more uncertain (and also model-dependent), and the concentration becomes more correlated with the other parameters.

Fitting the RIN has also caused the best-fit molar mass to fall by ~4%.

How much has this changed the systematic patterns in the residuals noted earlier (section 7)? Here are both line and gray scale residual plots for this revised fit:


Analysis1
Select graph type from list:
Residuals, line graph
Print
Copy
Add to Log
Graph Options
Single scan =
1
Every Nth scan, N=
4
Vary scan colors
Show legend
Spread residuals
overall r.m.s. deviation = 3.28E-3
Y(exp) - Y(fit)
radius (cm)
Log
Back
Next



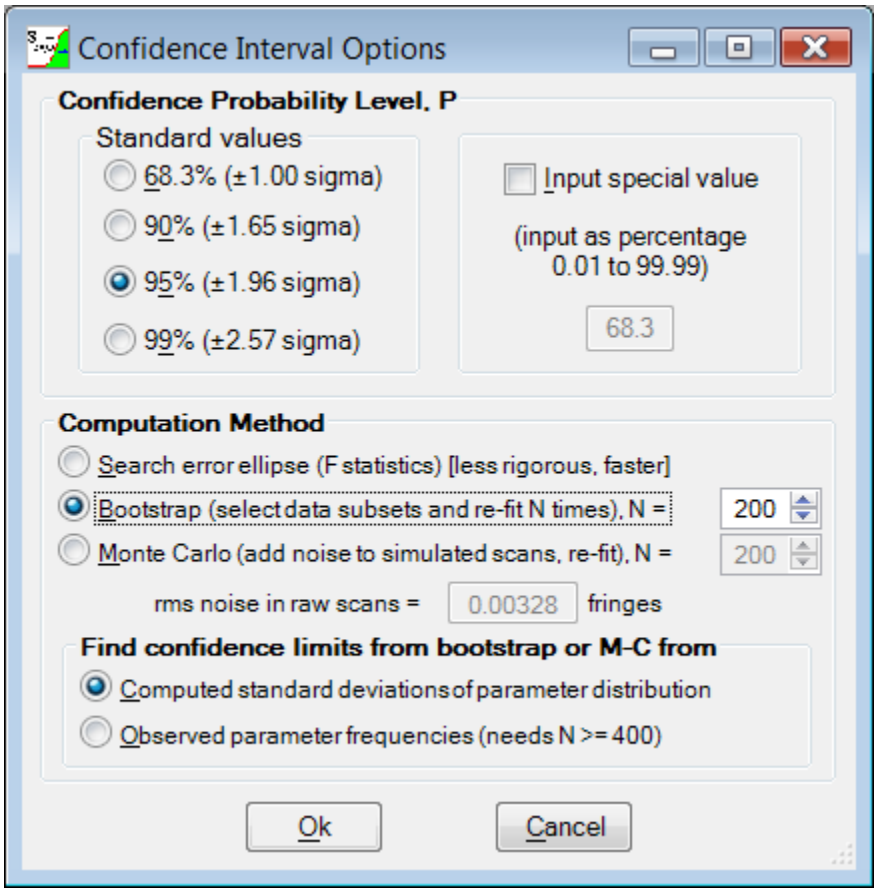
Clearly the magnitude of the systematic deviations has been significantly reduced, and the diagonal features in the contour plot are much less pronounced. On the other hand, the residuals are still far from random, and we still see runs of positive and negative residuals that persist over substantial portions of the cell, and often in a correlated fashion over several scans in a row (producing the horizontal bands evident in the graph above).

Overall though this fit is about as good as can be expected, and there is no clear reason to reject it and seek a more complex model. That is, we are done with fitting!

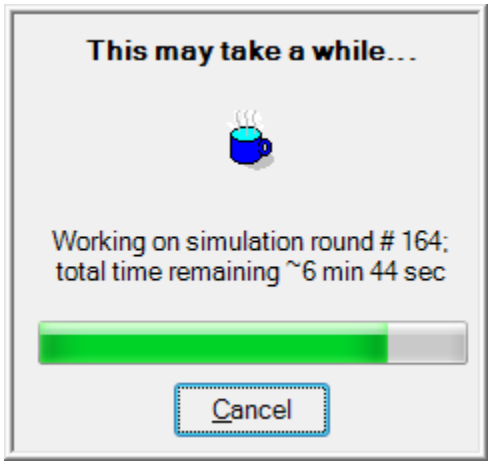
9. Calculating confidence limits

Now that we have a good fit, we will probably want a more precise estimate of the precision of the fitted parameters (and we definitely would if this result were going to be published). This is done by pressing the **Compute confidence intervals button (Section 3.2.6.4)** on the **Report fit results page (Section 3.2.6)**.

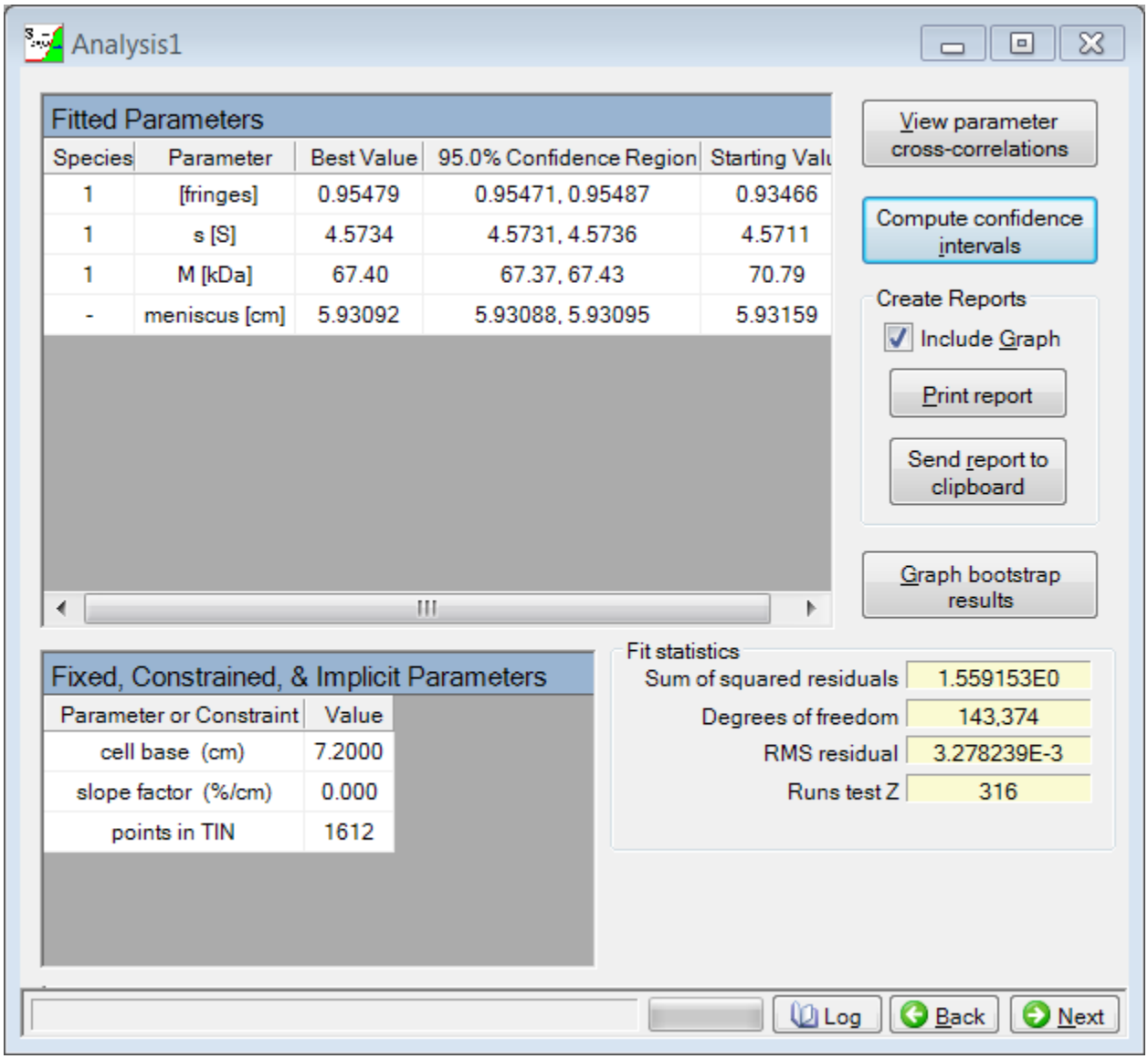
This brings up the following dialog box, which allows you to set the desired confidence level (for example 95% confidence or 68.3% confidence), and to choose among 3 alternate statistical approaches. More details about the various options can be found **here (Section 3.2.6.4)**.



In this case we will calculate the 95% confidence interval using 200 rounds of the bootstrap method. After clicking the Ok button the following dialog will be displayed showing the progress of the computations and the approximate amount of time remaining to complete them. For this fit ~25 min is required on a 2 GHz processor.



Below we see the **Report fit results page (Section 3.2.6)** after the calculations complete (our final result!).



Clearly this result implies that all of the parameters are very precisely determined. While that precision is correct statistically, this estimate does not necessarily reflect the actual precision (reproducibility) that would be observed if this experiment were repeated many times (because it cannot account for the impact of systematic errors). In actual practice when the signal/noise ratio of the raw data is high it will usually be the systematic noise that limits the precision, not the photometric noise of the detectors. In particular, the true precision of the sedimentation coefficients will usually be limited by systematic errors (such as temperature fluctuations and the absolute accuracy of the meniscus position) and thus will usually be significantly worse than the statistical estimate. **If you really want to know the precision of your results there is no shortcut---you must repeat the experiment multiple times.**

10. Revealing more about this data set

Now that our analysis is complete, it is time to reveal that we have actually been working with a simulated data set for BSA, but one where real noise (interference scans of an empty cell) has been added to the noise-free simulation. Thus in this case the correct values for the parameters (the values that went into the simulation) are known. The following table compares the true values to the final result from section 9 above:

Parameter	True Value	Best-fit Value [95% confidence interval]	Error (%)
Loading concentration (fringes)	0.960	0.95479 [0.95471, 0.95487]	-0.54
Sedimentation coefficient (S)	4.580	4.5734 [4.5731, 4.5736]	-0.14
Molar mass (kDa)	66.470	67.40 [67.37, 67.43]	1.40
Meniscus position (cm)	5.930	5.93092 [5.93088, 5.93095]	0.016

Unfortunately this shows that none of the fitted values are actually accurate (the true value does not fall within the 95% confidence interval). However this is not the fault of the data analysis, it is a consequence of the systematic noise present in the interference scans.

To prove that assertion, here are the results from the equivalent fit to the same simulation, but this time the noise added to the simulation was random (with a rms value corresponding to the rms residual of the fit in section 9 above).

Parameter	True Value	Best-fit Value [95% confidence interval]	Error (%)
Loading concentration (fringes)	0.960	0.96017 [0.96012, 0.96022]	0.018
Sedimentation coefficient (S)	4.580	4.57982 [4.57965, 4.57998]	-0.0039
Molar mass (kDa)	66.470	66.53 [66.51, 66.56]	0.090
Meniscus position (cm)	5.930	5.930004 [5.929987, 5.930019]	0.000067

Although even with random noise the 95% confidence interval does not include the correct value for most of the parameters, the errors are smaller than the accuracy attainable for any real experiment. These small errors are mostly related to the accuracy of the Lamm equation solutions used in generating the simulation and in the fitting function.

2.2 Tutorial: Multi-species fits

nIntroduction

Multi-species fits are not necessarily significantly different or more complex then single-species fits. When there aren't too many species, each of them is present at fractions above ~10%, and the experiment has a good signal/noise ratio, you should have no difficulty provided you can make reasonable initial guesses for the sedimentation coefficients and molar mass of each species. Those initial guesses could potentially come from a *c(s)* analysis using SEDFIT (see **Tutorial: Using a c(s) distribution with SVEDBERG (Section 2.3)**).

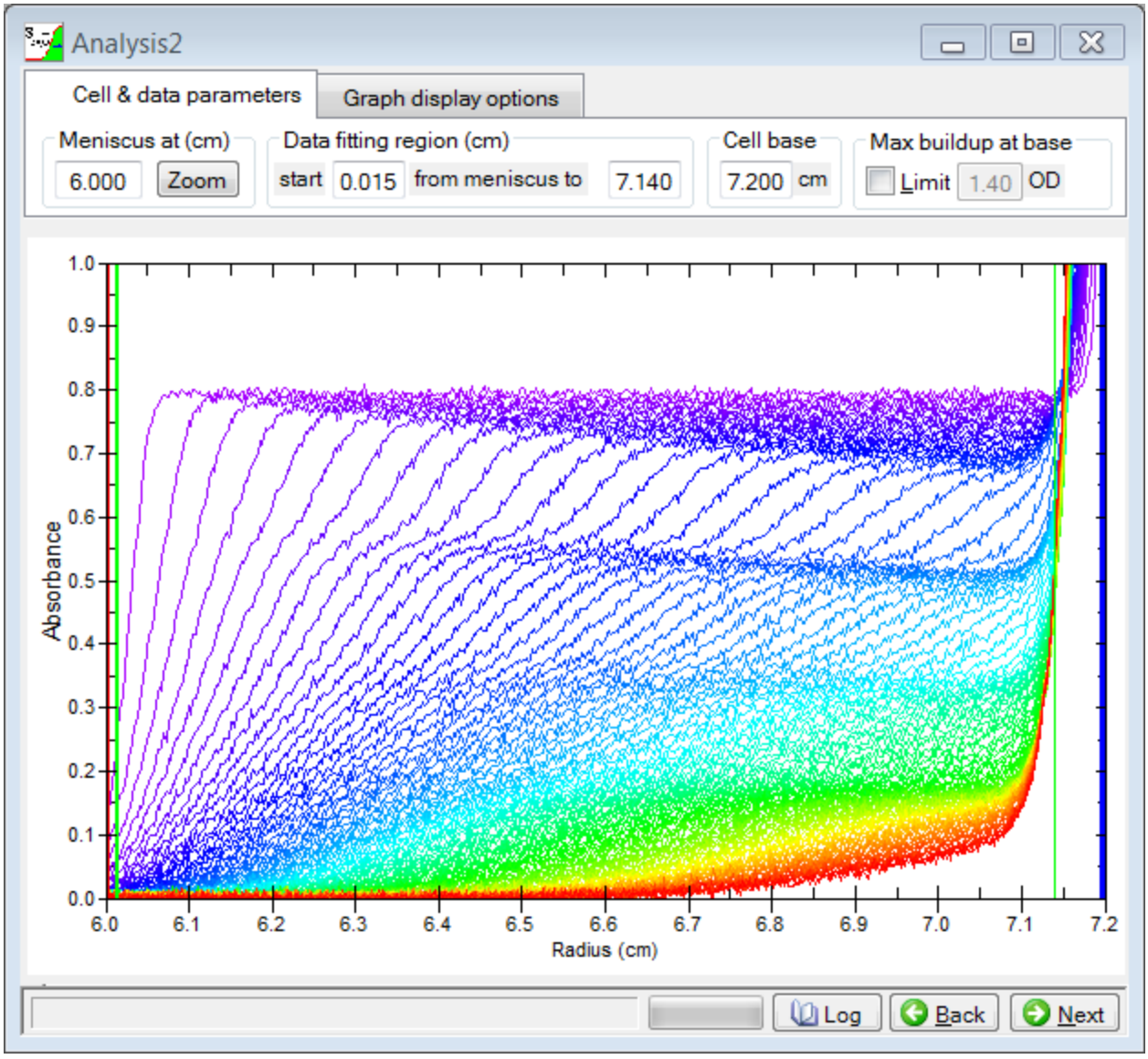
The more difficult situations arise when the species are poorly resolved, or when some species are present at fractions of only a few percent or less, or the signal/noise ratio is poor. In such situations the raw data may simply not contain enough information to uniquely determine the molar masses of all species (especially those present at low levels). Thus it sometimes is necessary to input information obtained from other techniques (for example, holding a molar mass value fixed at the known sequence mass for a component of the sample). Another useful (and common) approach is to constrain the masses based on reasonable assumptions. Two common such assumptions are:

1. assuming all species have similar f/f_0 ratios (the assumption used by the *c(s)* method in SEDFIT)
2. assuming they represent different oligomers of the same monomer and therefore their masses are related by integer ratios

The examples below illustrate both an 'easy' sample and a 'difficult' one requiring assumptions about the properties of the sample.

An easy case: an equal mixture of 4 proteins

Our first example will be fitting simulated data for a sample containing equal parts of 4 proteins: a 20 kDa protein at 2 S (f/f_0 = 1.32), a 50 kDa protein at 4 S (f/f_0 = 1.22), a 150 kDa protein at 6 S (f/f_0 = 1.69), and a 300 kDa protein at 9 S (f/f_0 = 1.79). This simulation of an absorbance experiment at 55,000 rpm and with scans recorded every 4 min (100 total scans) assumes a total loading concentration of 0.8 OD and an rms noise level of 0.005 OD. The **Set meniscus & data region page (Section 3.2.2)** below shows the simulated data and the fitting regions we will use.



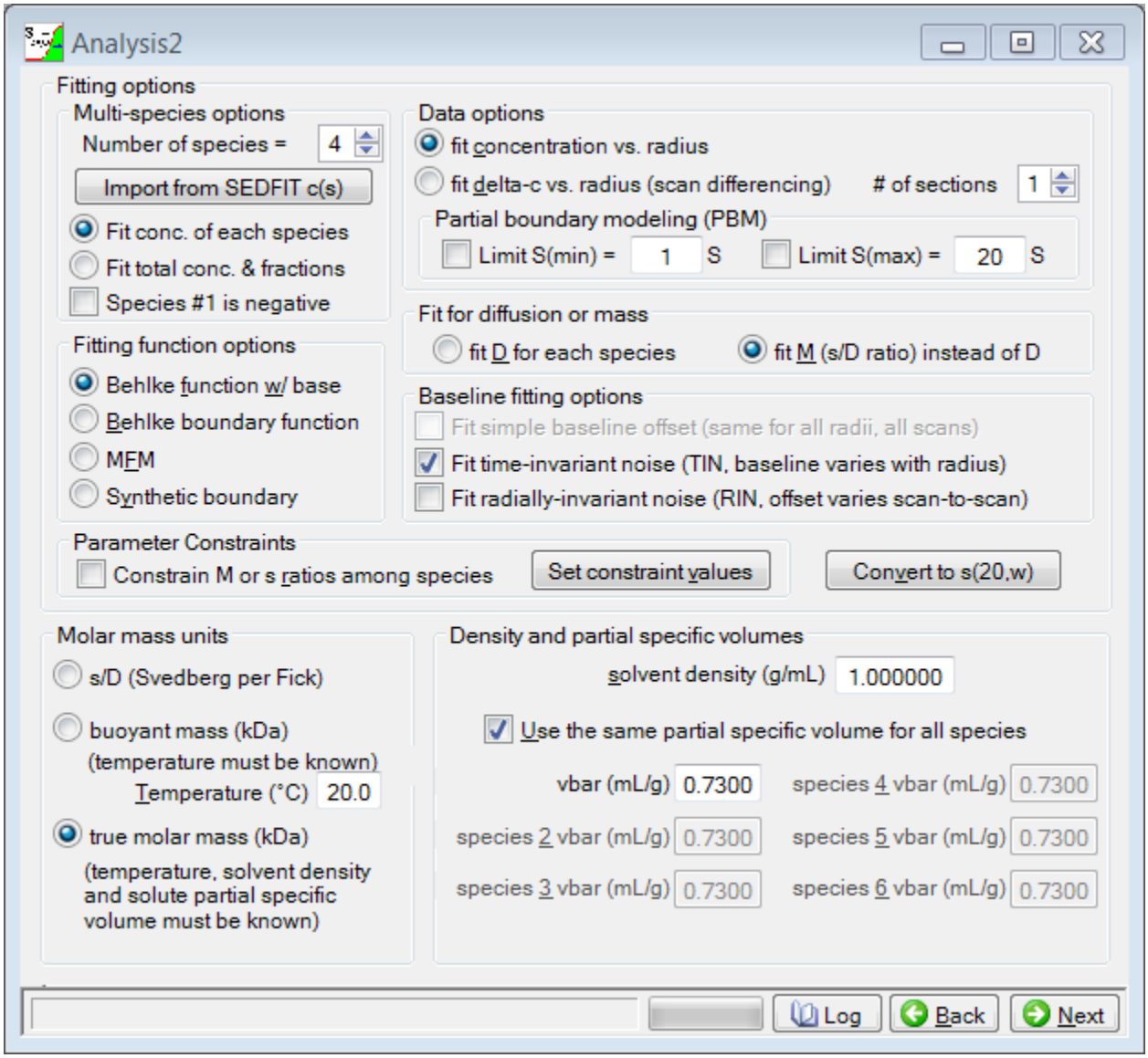
Moving forward to the **Select fitting model page (Section 3.2.3)** using the the **Next button (C:\Users\Public\Documents\Visual Studio Projects\Svedberg Help Source\{46451615-072B-4FCC-B31F-82953CB24396})**, all we need to do is:

1. Use the up/down spinner in the upper left corner to increase the number of species to 4
2. Optionally click the **Fit total concentration and fractions radio button** (upper left). Whether it is best to fit for the loading concentration of each species versus total & species fractions primarily depends on

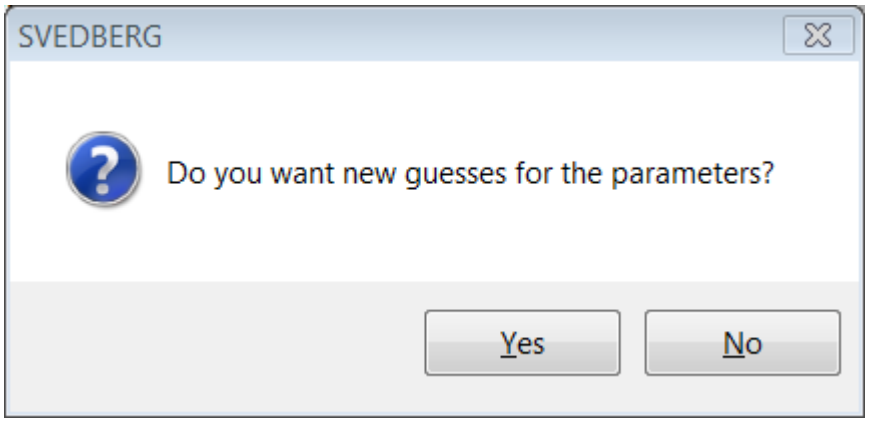
which quantity interests you the most. Obviously it is always possible to manually calculate fractions from individual concentrations or *vice versa*, but it is not possible to convert the confidence intervals without re-fitting because the contributions of different species cross-correlate in different ways for these two options. For this example we will compute individual concentrations.

3. Click on the **Fit time invariant noise checkbox** in **Baseline fitting options (Section 3.2.3.5)** (even though this is simulated data we will allow for a varying baseline profile since that would likely be true for real experiments)
4. Set the solvent density and partial specific volume(s) to the appropriate values. (In this case the simulation used a density of 1.0 g/mL and vbar = 0.73 mL/g)

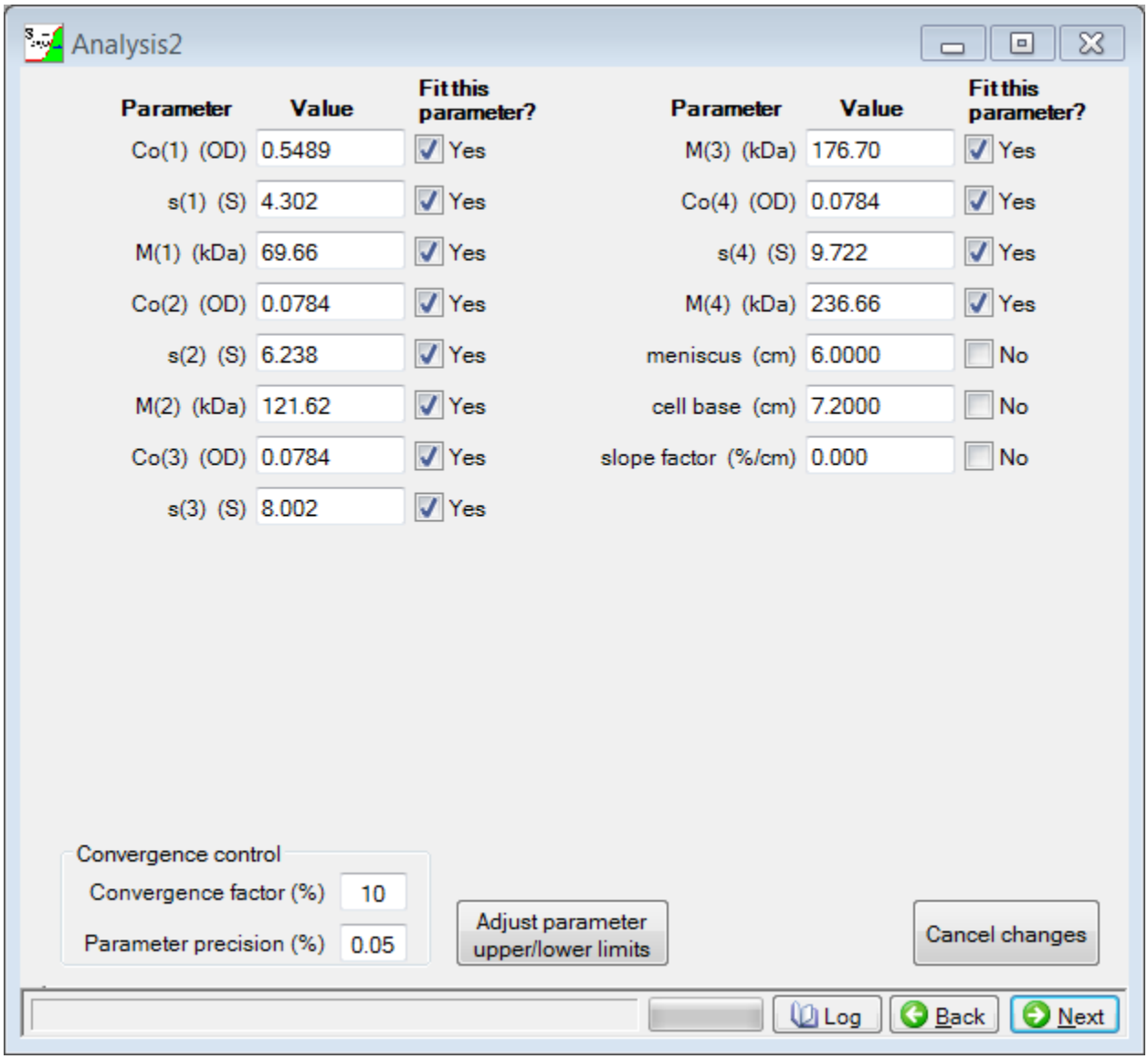
This gives us a fit model that looks like this:


Analysis2
Fitting options
Multi-species options
Number of species = 4
Import from SEDFIT c(s)
Fit conc. of each species
Fit total conc. & fractions
Species #1 is negative
Fitting function options
Behlke function w/ base
Behlke boundary function
MEM
Synthetic boundary
Data options
fit concentration vs. radius
fit delta-c vs. radius (scan differencing)
# of sections 1
Partial boundary modeling (PBM)
Limit S(min) = 1 S
Limit S(max) = 20 S
Fit for diffusion or mass
fit D for each species
fit M (s/D ratio) instead of D
Baseline fitting options
Fit simple baseline offset (same for all radii, all scans)
Fit time-invariant noise (TIN, baseline varies with radius)
Fit radially-invariant noise (RIN, offset varies scan-to-scan)
Parameter Constraints
Constrain M or s ratios among species
Set constraint values
Convert to s(20,w)
Molar mass units
s/D (Svedberg per Fick)
buoyant mass (kDa)
(temperature must be known)
Temperature (°C) 20.0
true molar mass (kDa)
(temperature, solvent density and solute partial specific volume must be known)
Density and partial specific volumes
solvent density (g/mL) 1.000000
Use the same partial specific volume for all species
vbar (mL/g) 0.7300
species 4 vbar (mL/g) 0.7300
species 2 vbar (mL/g) 0.7300
species 5 vbar (mL/g) 0.7300
species 3 vbar (mL/g) 0.7300
species 6 vbar (mL/g) 0.7300
Log
Back
Next

When we move forward to the **Set/alter fitting parameters page (Section 3.2.4)** to enter initial guesses we see this dialog (and will answer Yes):



That gives us a set of initial guesses that looks like this:



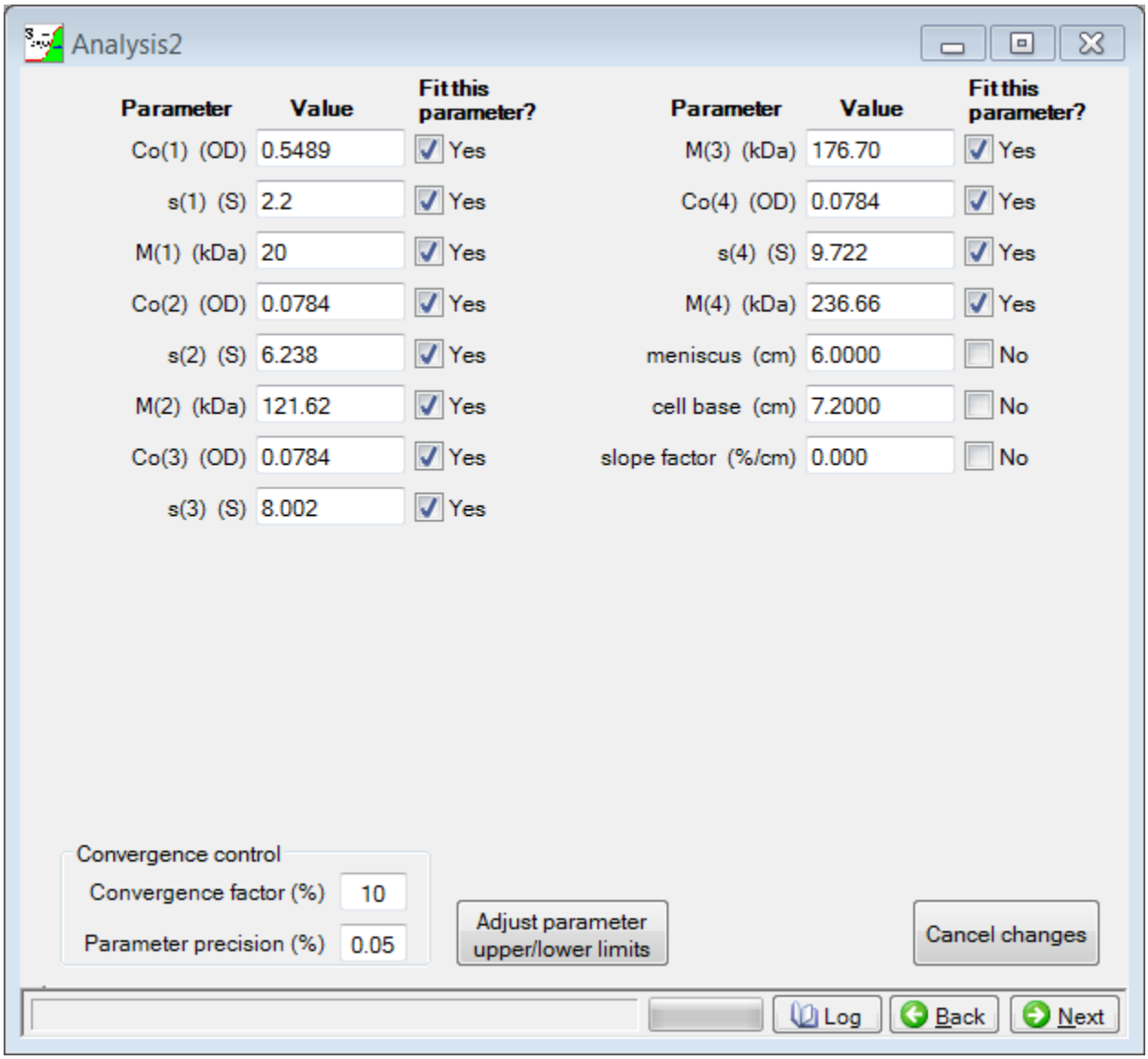
The program wizard of course knows nothing about the nature of the sample. For multi-species fits it does the following:

1. it assigns one tenth of the total to each of species 2 to N, and the balance to species 1
2. it assigns the sedimentation coefficient of species 1 as 0.9 times the overall average and then values for species 2 to N by assuming they are dimer, trimer, tetramer, hexamer, and octamer of species 1
3. it enters an initial guess for the molar mass of each species based on its sedimentation coefficient and an assumed frictional coefficient ratio (f/f_0 ratio) of 1.4

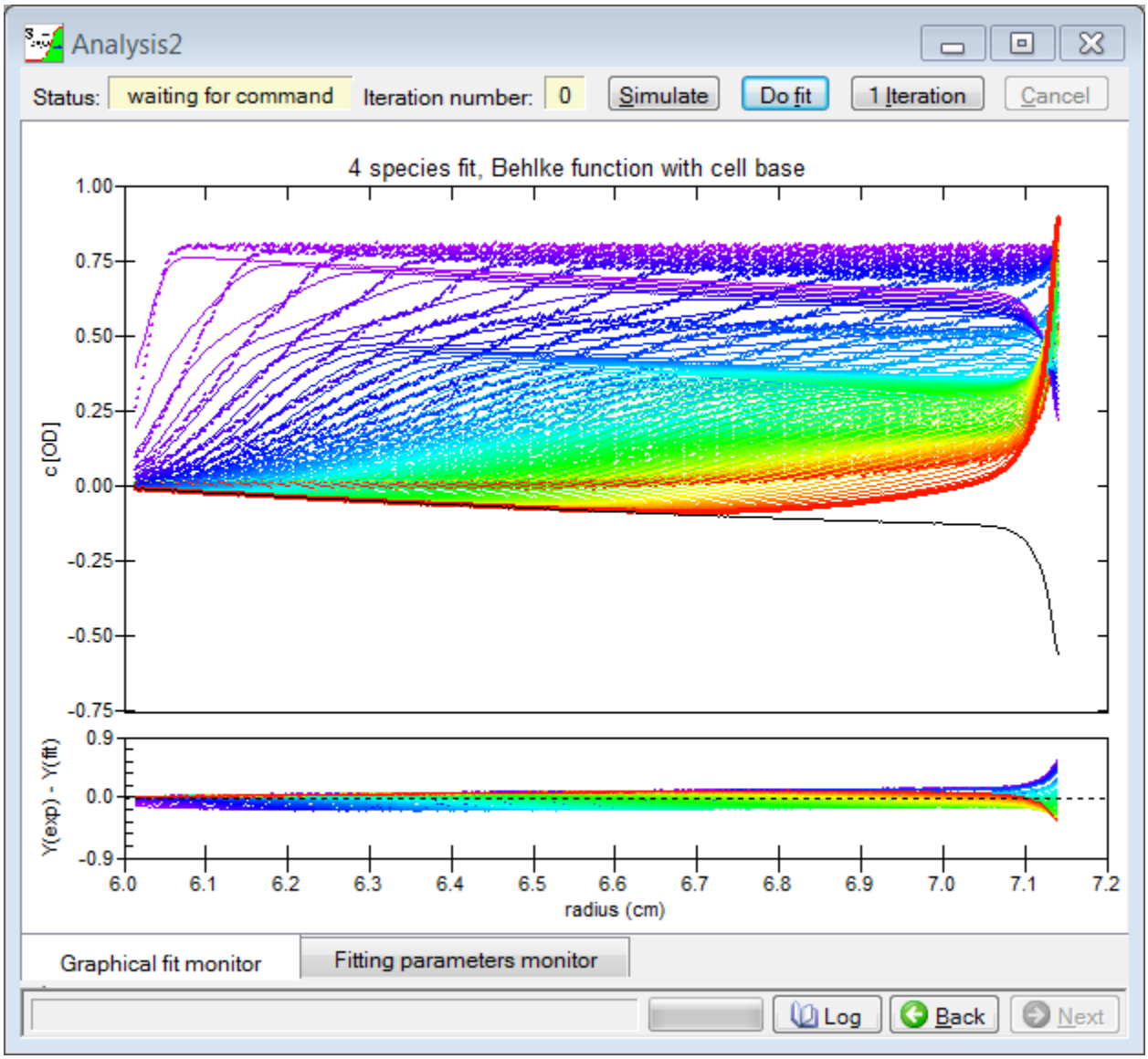
In this case the wizard's guesses of the sedimentation coefficients are poor (this is far more important than its incorrect guess about the species concentrations), and the sedimentation coefficients are too far off for the fit to be able to converge with so many species. However if we simply new that one of the species had a mass of ~20 kDa, and therefore guessed it has a f/f_0 ratio of 1.2 (appropriate for a hydrated compact globular protein) and thus a sedimentation coefficient of 2.2, changing those initial guesses for species 1 will be sufficient to get the fit to converge.

Note that for this simulated data we know the correct meniscus position is 6.00 cm. For a real experiment you might want to fit the meniscus position (see **Should I fit the meniscus position? (Section 5.1.1)**). However if you wanted to fit the meniscus position, given that the initial guesses are probably going to be poor it would be best to initially fix the meniscus position, and then later allow it to vary (after the parameters are better optimized).

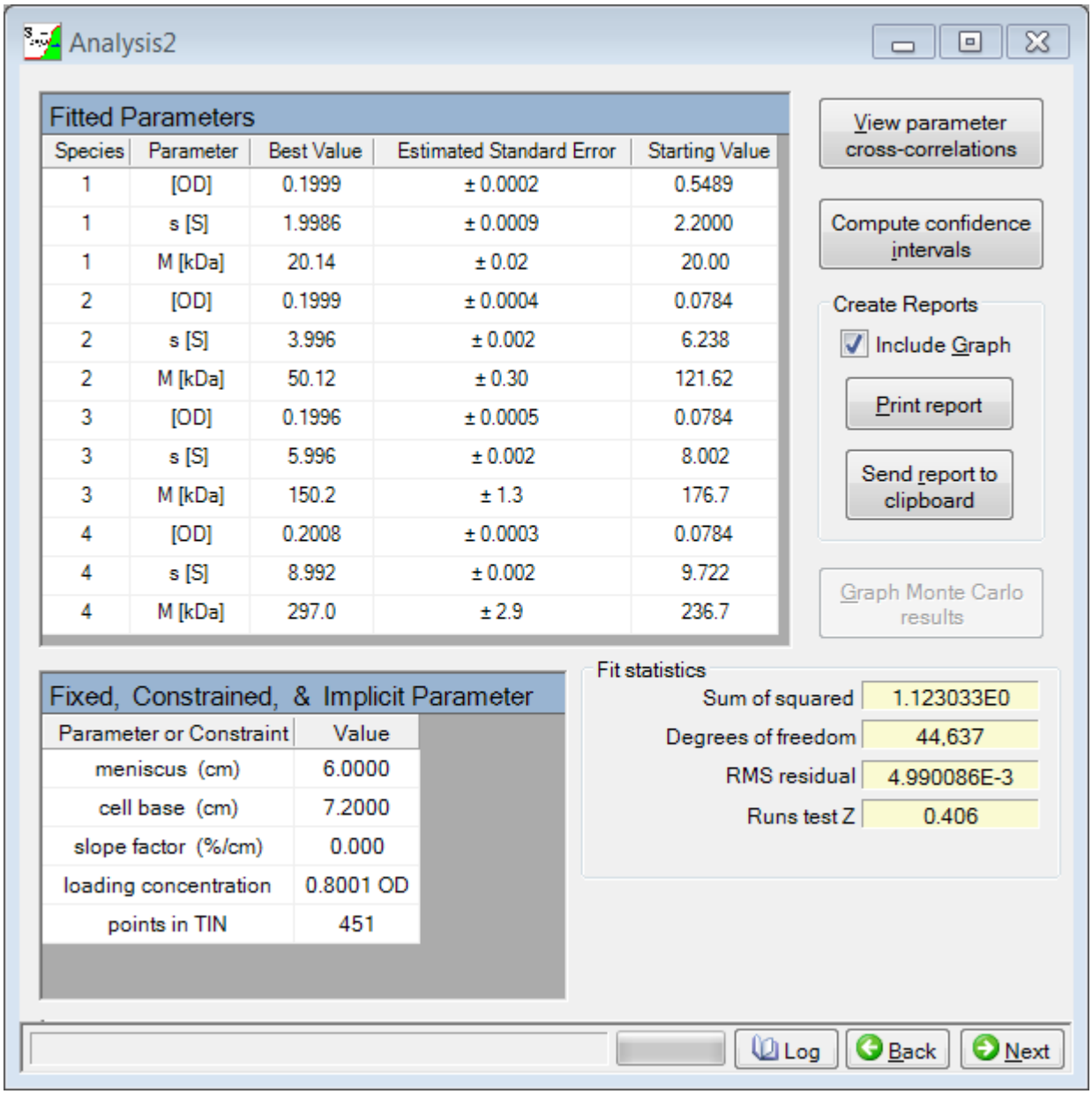
Thus our fit setup and initial guesses look like this:



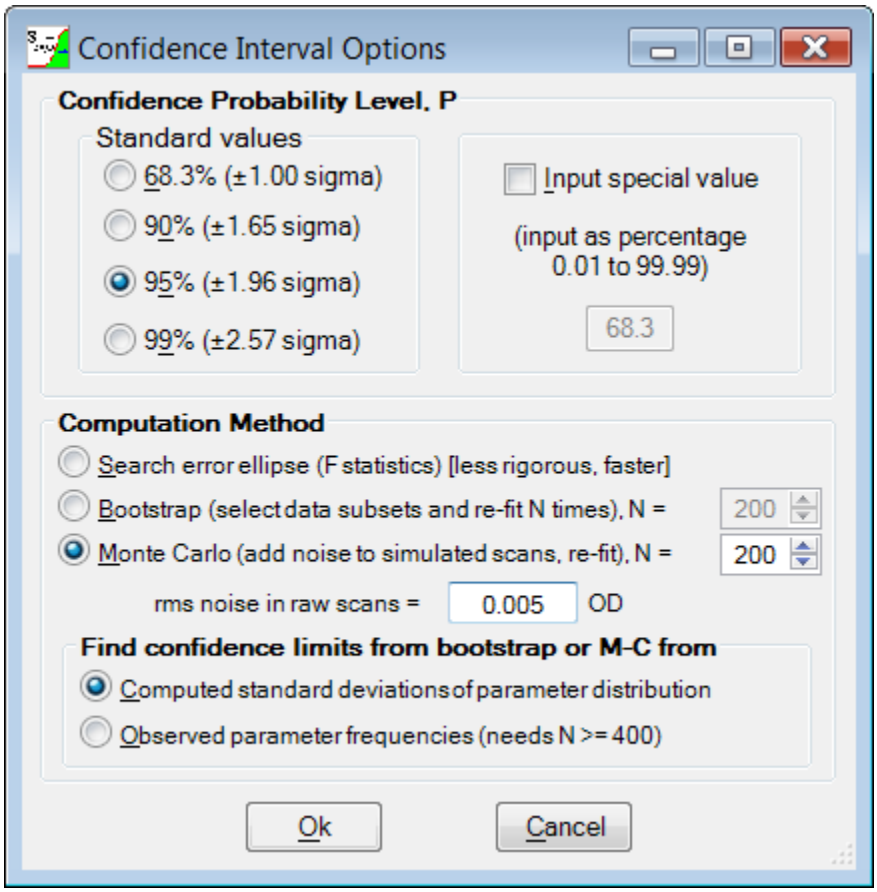
Moving forward to the **Perform least-squares fit page (Section 3.2.5)**, we see that these guesses give a rather poor fit, but we will proceed with the fit nonetheless.



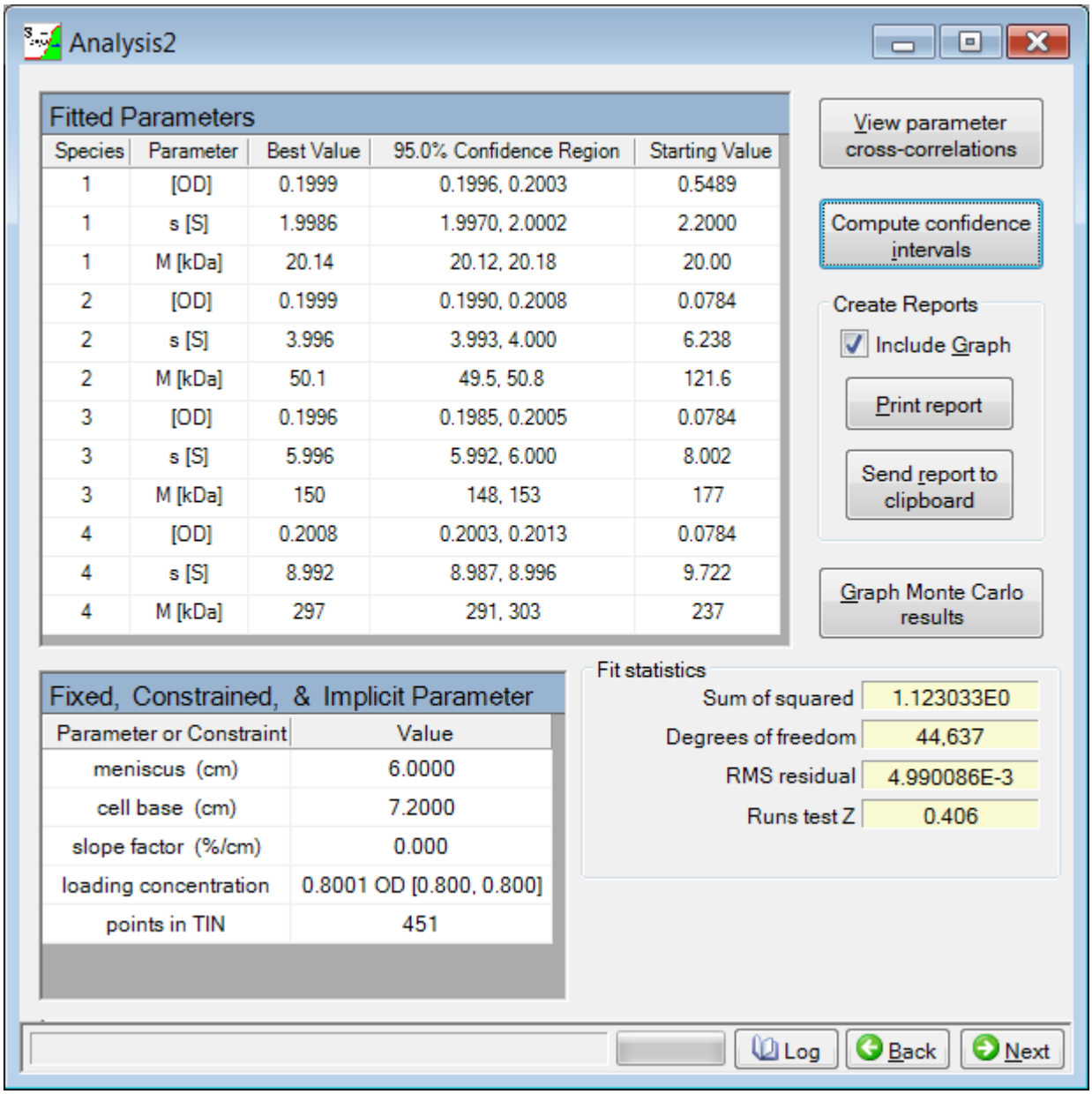
During the first ~15 iterations of the fit the fitter struggles a bit to move toward a global minimum, but once it drops the sedimentation coefficient of species 2 into the right range the fit converges fairly quickly and is complete after 21 iterations (taking ~55 sec total on a 2.2 GHz CPU), with the results shown below.


Analysis2
Fitted Parameters
Species Parameter Best Value Estimated Standard Error Starting Value
1 [OD] 0.1999 ± 0.0002 0.5489
1 s [S] 1.9986 ± 0.0009 2.2000
1 M [kDa] 20.14 ± 0.02 20.00
2 [OD] 0.1999 ± 0.0004 0.0784
2 s [S] 3.996 ± 0.002 6.238
2 M [kDa] 50.12 ± 0.30 121.62
3 [OD] 0.1996 ± 0.0005 0.0784
3 s [S] 5.996 ± 0.002 8.002
3 M [kDa] 150.2 ± 1.3 176.7
4 [OD] 0.2008 ± 0.0003 0.0784
4 s [S] 8.992 ± 0.002 9.722
4 M [kDa] 297.0 ± 2.9 236.7
View parameter cross-correlations
Compute confidence intervals
Create Reports
Include Graph
Print report
Send report to clipboard
Graph Monte Carlo results
Fixed, Constrained, & Implicit Parameter
Parameter or Constraint Value
meniscus (cm) 6.0000
cell base (cm) 7.2000
slope factor (%/cm) 0.000
loading concentration 0.8001 OD
points in TIN 451
Fit statistics
Sum of squared 1.123033E0
Degrees of freedom 44,637
RMS residual 4.990086E-3
Runs test Z 0.406
Log
Back
Next

Clearly the best-fit values are very close to the inputs to the simulation (the accuracy is good). To better evaluate the precision of the results (can we get all the properties of all 4 species with good precision from data with this signal/noise level?) we need a rigorous computation of the confidence levels. Clicking the **Compute confidence intervals** button brings up a dialog box where we will select a 95% confidence level and 200 rounds of Monte Carlo simulation with 0.005 OD rms noise (the known noise level of these data), as shown below.

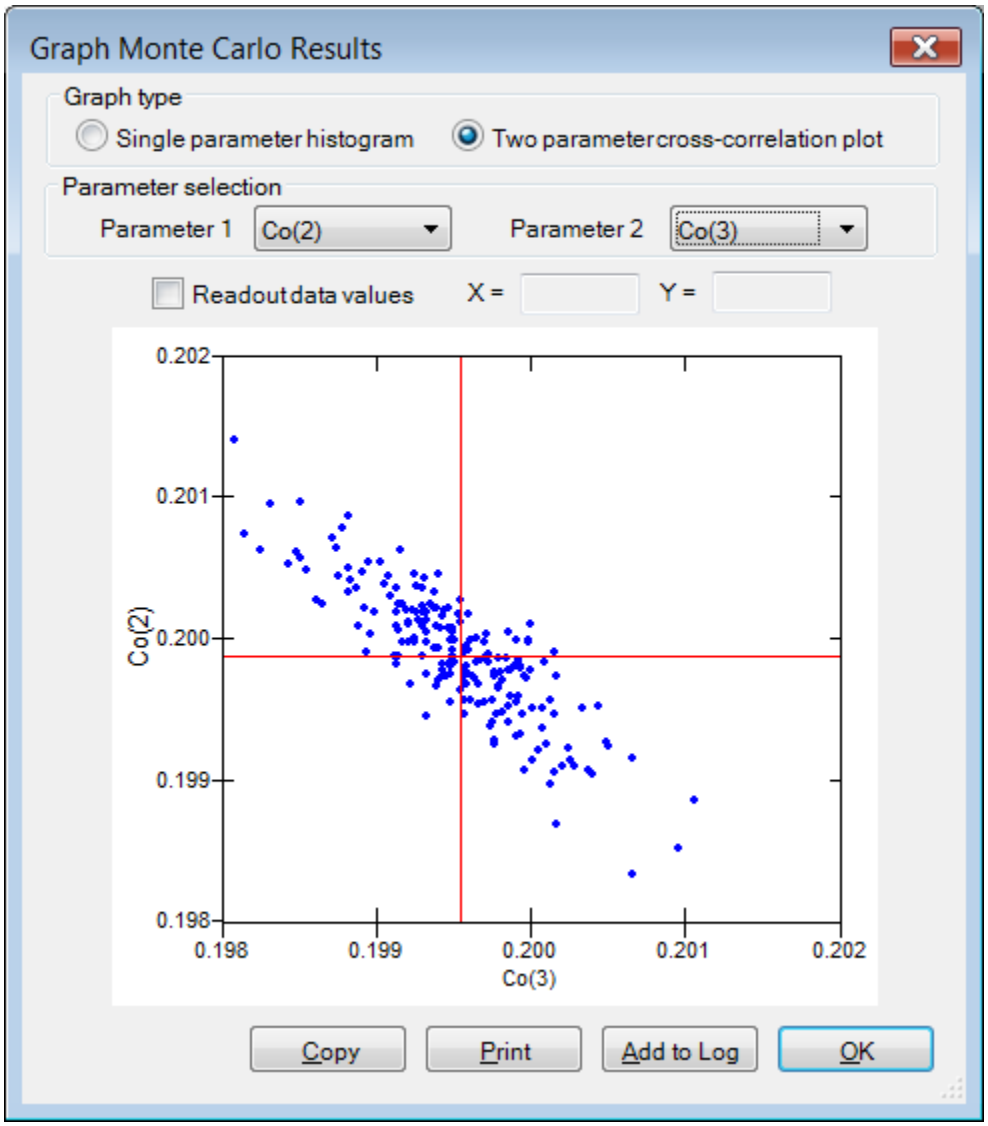


This gives the final results shown below.


Analysis2
Fitted Parameters
Species | Parameter | Best Value | 95.0% Confidence Region | Starting Value
1 | [OD] | 0.1999 | 0.1996, 0.2003 | 0.5489
1 | s [S] | 1.9986 | 1.9970, 2.0002 | 2.2000
1 | M [kDa] | 20.14 | 20.12, 20.18 | 20.00
2 | [OD] | 0.1999 | 0.1990, 0.2008 | 0.0784
2 | s [S] | 3.996 | 3.993, 4.000 | 6.238
2 | M [kDa] | 50.1 | 49.5, 50.8 | 121.6
3 | [OD] | 0.1996 | 0.1985, 0.2005 | 0.0784
3 | s [S] | 5.996 | 5.992, 6.000 | 8.002
3 | M [kDa] | 150 | 148, 153 | 177
4 | [OD] | 0.2008 | 0.2003, 0.2013 | 0.0784
4 | s [S] | 8.992 | 8.987, 8.996 | 9.722
4 | M [kDa] | 297 | 291, 303 | 237
View parameter cross-correlations
Compute confidence intervals
Create Reports
Include Graph
Print report
Send report to clipboard
Graph Monte Carlo results
Fixed, Constrained, & Implicit Parameter
Parameter or Constraint | Value
meniscus (cm) | 6.0000
cell base (cm) | 7.2000
slope factor (%/cm) | 0.000
loading concentration | 0.8001 OD [0.800, 0.800]
points in TIN | 451
Fit statistics
Sum of squared | 1.123033E0
Degrees of freedom | 44,637
RMS residual | 4.990086E-3
Runs test Z | 0.406
Log
Back
Next

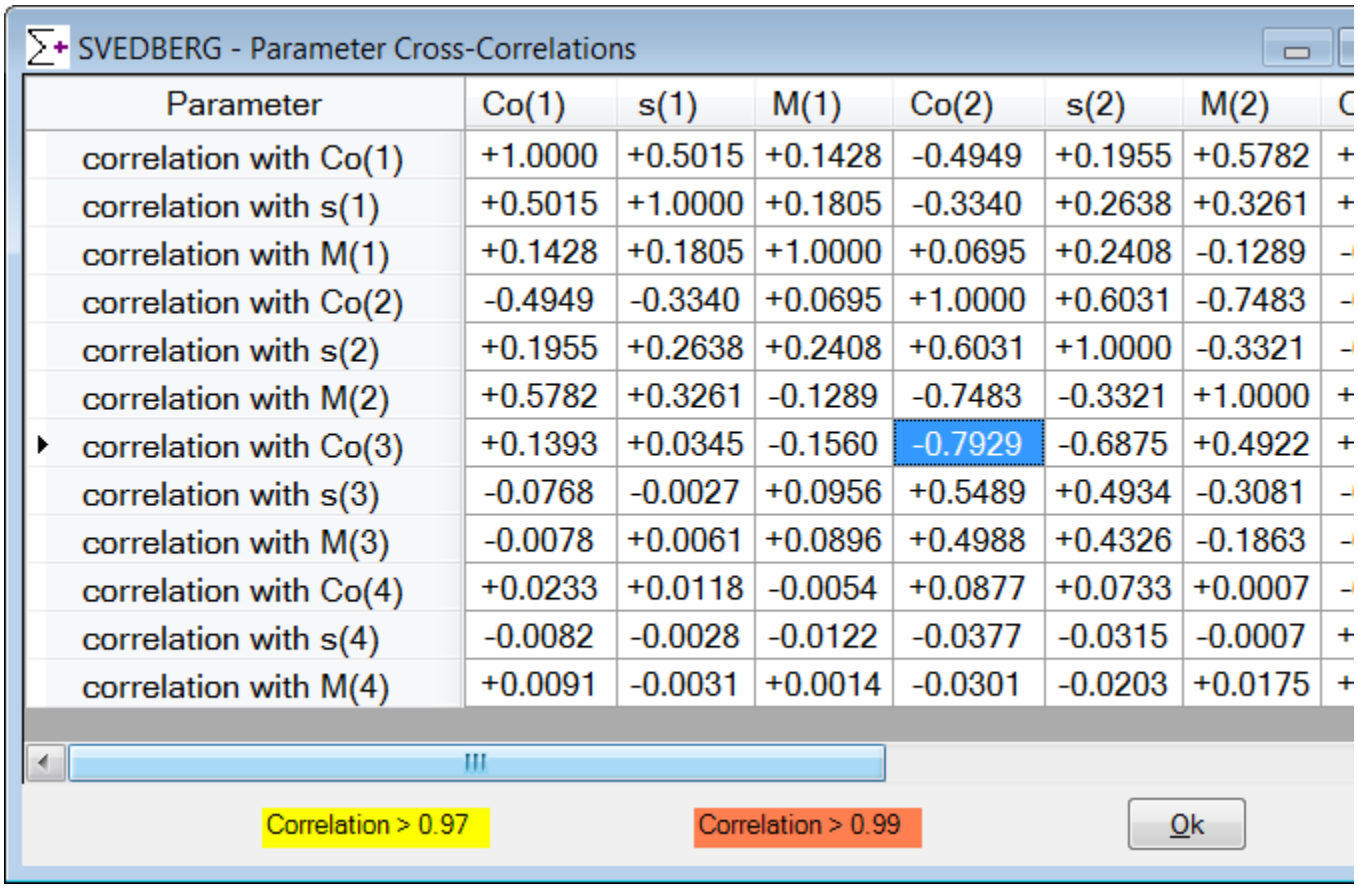
This shows that the concentration, sedimentation coefficient, and molar mass of all 4 components can be determined with a precision (95% confidence range) of ±2% or better, and an accuracy (difference between the best-fit value and the value that went into the simulation) of ±1% or better. These are, of course, simulated data, but the assumed noise levels are realistic for a properly-functioning XL-A.

Another useful way to look at the degree to which the properties of the 4 species can be separated (deconvoluted from the raw data) is to make a cross-correlation plot between the values of two parameters for the entire set of Monte Carlo results. By pushing the **Graph Monte Carlo results** button, clicking the **Two-parameter cross-correlation plot** radio button, and selecting the load concentration parameters for species 2 and 3 [the Co(2) and Co(3) parameters] from the drop-down list boxes, we get the following graph:



The fact that the parameters cluster within an ellipse along the diagonal, rather than distributing randomly around the best-fit values (marked by the red lines), shows that these two concentrations are fairly highly correlated (a negative correlation in this case). This means that these two concentrations cannot be determined entirely independently---an increase in the concentration of species 2 can be largely compensated by decreasing the concentration of species 3. This correlation is strongest for species 2 and 3 because their relative sedimentation coefficients are the closest (they are the most poorly physically separated).

This moderately high parameter cross-correlation can also be seen in the table obtained using the **View parameter cross-correlations** button, shown below.



Parameter	Co(1)	s(1)	M(1)	Co(2)	s(2)	M(2)
correlation with Co(1)	+1.0000	+0.5015	+0.1428	-0.4949	+0.1955	+0.5782
correlation with s(1)	+0.5015	+1.0000	+0.1805	-0.3340	+0.2638	+0.3261
correlation with M(1)	+0.1428	+0.1805	+1.0000	+0.0695	+0.2408	-0.1289
correlation with Co(2)	-0.4949	-0.3340	+0.0695	+1.0000	+0.6031	-0.7483
correlation with s(2)	+0.1955	+0.2638	+0.2408	+0.6031	+1.0000	-0.3321
correlation with M(2)	+0.5782	+0.3261	-0.1289	-0.7483	-0.3321	+1.0000
correlation with Co(3)	+0.1393	+0.0345	-0.1560	-0.7929	-0.6875	+0.4922
correlation with s(3)	-0.0768	-0.0027	+0.0956	+0.5489	+0.4934	-0.3081
correlation with M(3)	-0.0078	+0.0061	+0.0896	+0.4988	+0.4326	-0.1863
correlation with Co(4)	+0.0233	+0.0118	-0.0054	+0.0877	+0.0733	+0.0007
correlation with s(4)	-0.0082	-0.0028	-0.0122	-0.0377	-0.0315	-0.0007
correlation with M(4)	+0.0091	-0.0031	+0.0014	-0.0301	-0.0203	+0.0175

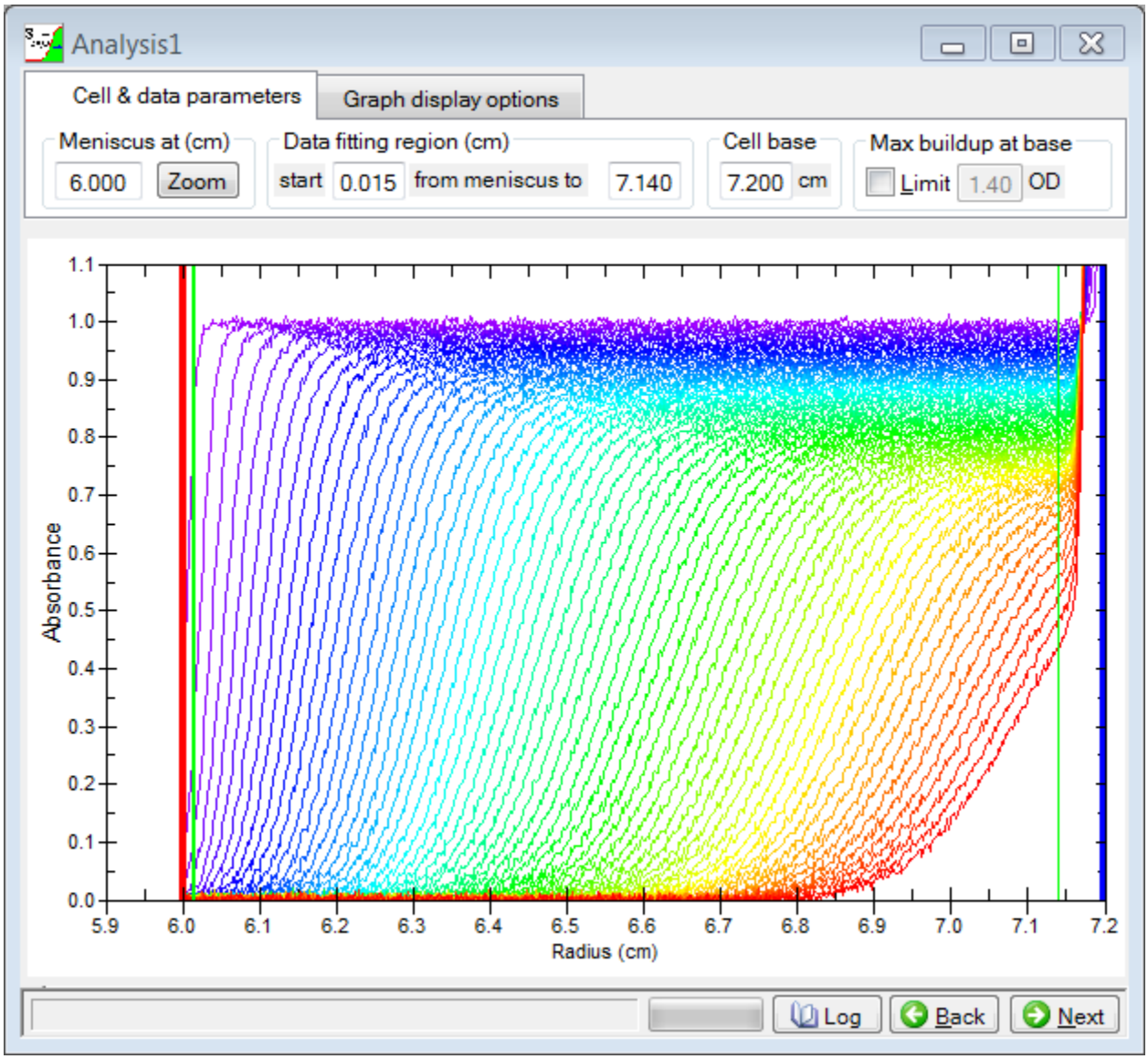
Note however that the values in this table are based on linear statistics, and the graphs are often more informative. Overall none of the parameter cross-correlations for this fit are high enough to cause major concerns (by default only values above 0.97 are flagged and generate warnings), but whenever there is reason to think you may be trying to extract too much information from the experiment one way to examine that is to look for high cross-correlations.

A more difficult example illustrating the use of parameter constraints

As mentioned in the Introduction, it is usually much more difficult to handle a sample containing minor components at very low levels. This example will use a simulation for a monoclonal antibody sample (6.2 S, 150 kDa) which contains 2% irreversible dimer (9.0 S), 1% trimer (11.6 S), and 0.5% tetramer (14.0 S), run at 40,000 rpm with scans every 4 min. The total loading concentration is 1 OD with 0.005 OD rms noise.

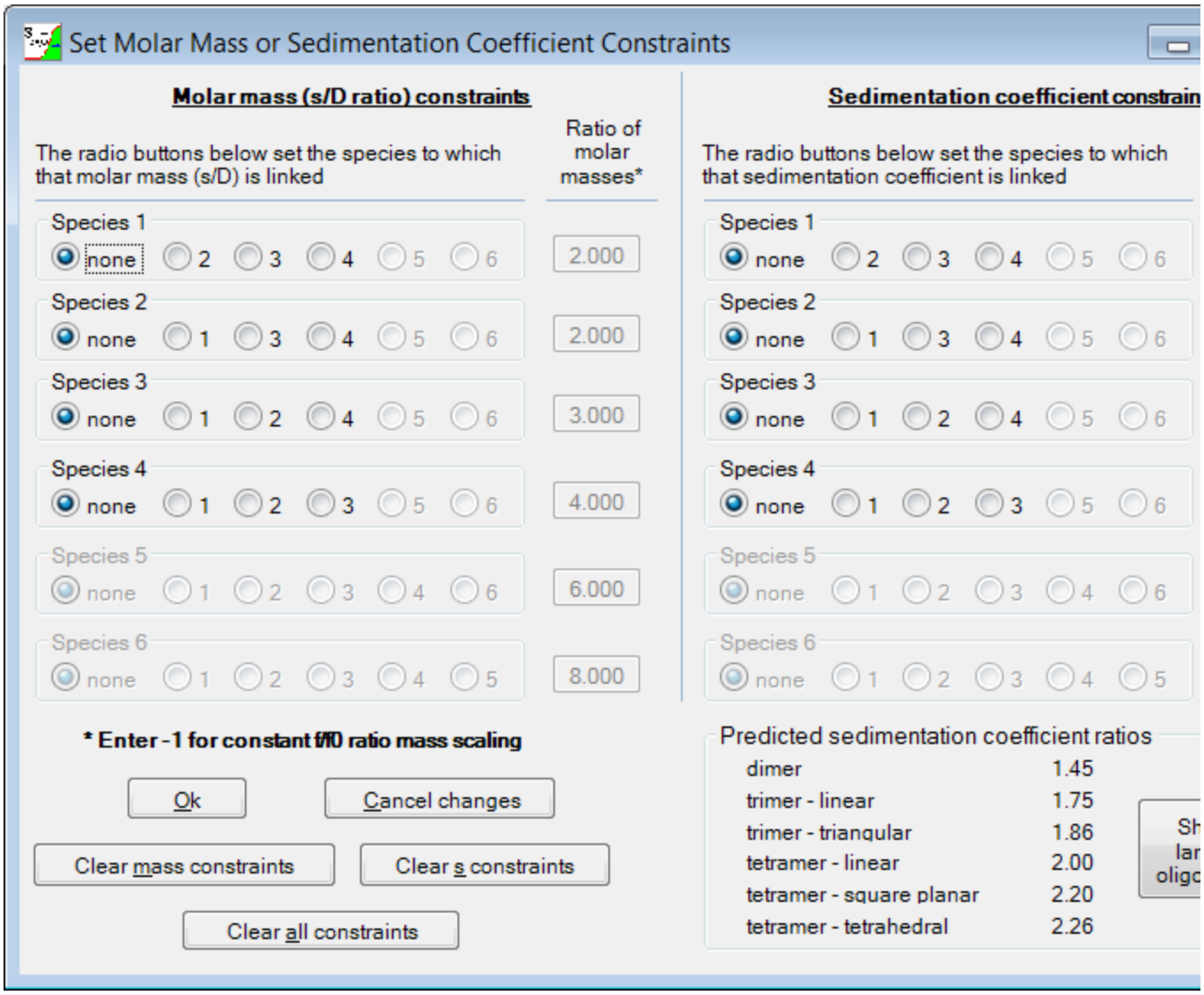
Let's suppose we knew that the sample contains small amounts of 3 aggregate species (for example from size exclusion chromatography or non-reducing SDS-PAGE), and we believe these are most likely dimer, trimer, and tetramer. What we would like to do is to quantitate the fractions of these aggregates, and to check whether their sedimentation coefficients are consistent with assignments as dimer, trimer, and tetramer.

After loading all the scans, here is how the raw data and the fitting region setup look:



Next we go to the **Select fitting model page (Section 3.2.3)** and do the following:

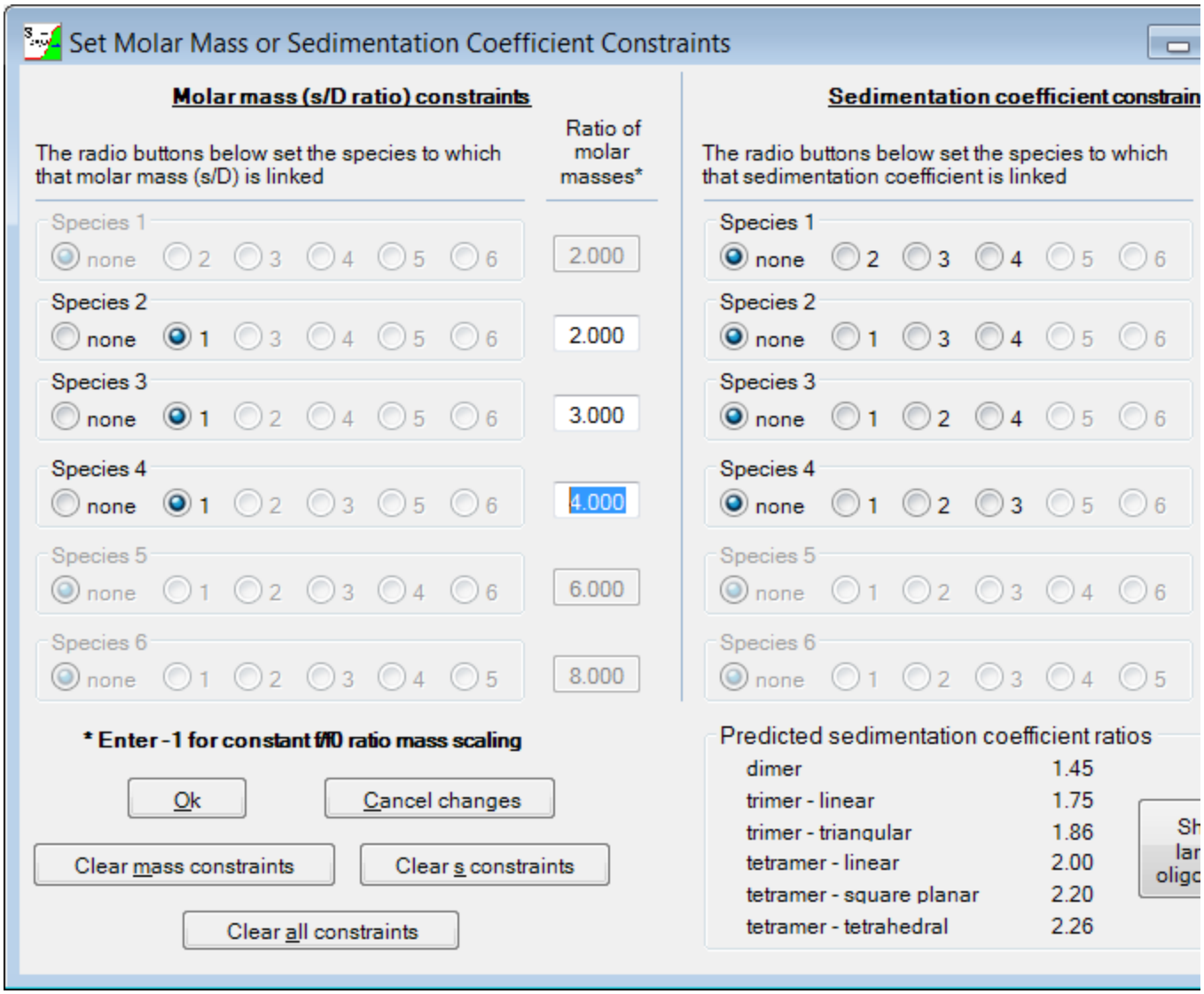
1. Use the up/down spinner in the upper left corner to increase the number of species to 4
2. Click the **Fit total concentration and fractions radio button** (upper left) to specify that we want to measure the species fractions.
3. Click on the **Fit time invariant noise checkbox** in **Baseline fitting options (Section 3.2.3.5)** (even though this is simulated data we will allow for a varying baseline profile since that would likely be true for real experiments)
4. Set the solvent density and partial specific volume(s) to the appropriate values. (In this case the simulation used a density of 1.0 g/mL and vbar = 0.73 mL/g)
5. Click on the **Set constraint values** button to bring up the mass and sedimentation coefficient constraints dialog


Set Molar Mass or Sedimentation Coefficient Constraints
Molar mass (s/D ratio) constraints
The radio buttons below set the species to which that molar mass (s/D) is linked
Ratio of molar masses*
Species 1 none 2 3 4 5 6 2.000
Species 2 none 1 3 4 5 6 2.000
Species 3 none 1 2 4 5 6 3.000
Species 4 none 1 2 3 5 6 4.000
Species 5 none 1 2 3 4 6 6.000
Species 6 none 1 2 3 4 5 8.000
* Enter -1 for constant f/f0 ratio mass scaling
Ok
Cancel changes
Clear mass constraints
Clear s constraints
Clear all constraints
Sedimentation coefficient constrain
The radio buttons below set the species to which that sedimentation coefficient is linked
Species 1 none 2 3 4 5 6
Species 2 none 1 3 4 5 6
Species 3 none 1 2 4 5 6
Species 4 none 1 2 3 5 6
Species 5 none 1 2 3 4 6
Species 6 none 1 2 3 4 5
Predicted sedimentation coefficient ratios
dimer 1.45
trimer - linear 1.75
trimer - triangular 1.86
tetramer - linear 2.00
tetramer - square planar 2.20
tetramer - tetrahedral 2.26

Given what we know about the sample, we are (at least initially) going to constrain the molar masses of the aggregate species to be integer multiples of the (unknown) monomer mass. To do this we need to link those species to the monomer. At this point no species properties have been defined, so the species numbering is arbitrary, but typically the main species is designated as species #1, so for this fit the monomer will be #1.

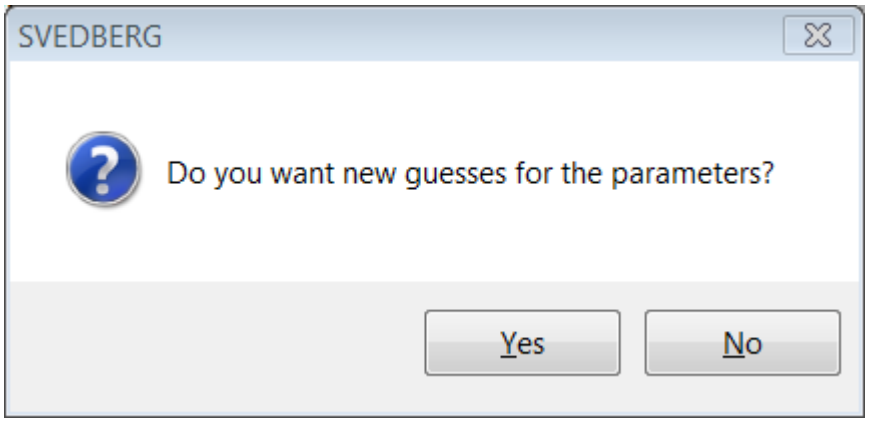
The links are formed by clicking on the radio buttons labeled 1 within the groups labeled Species 2, 3, and 4 on the left half of this dialog box, which links the molar mass of species 2-4 to that of species 1. Doing that also activates the **Ratio of molar masses** text boxes, where you should enter the integer ratio corresponding to the stoichiometry. In this case the default ratio values for species 2-4 already correspond to dimer, trimer, and tetramer.

At this point the dialog box looks like this:

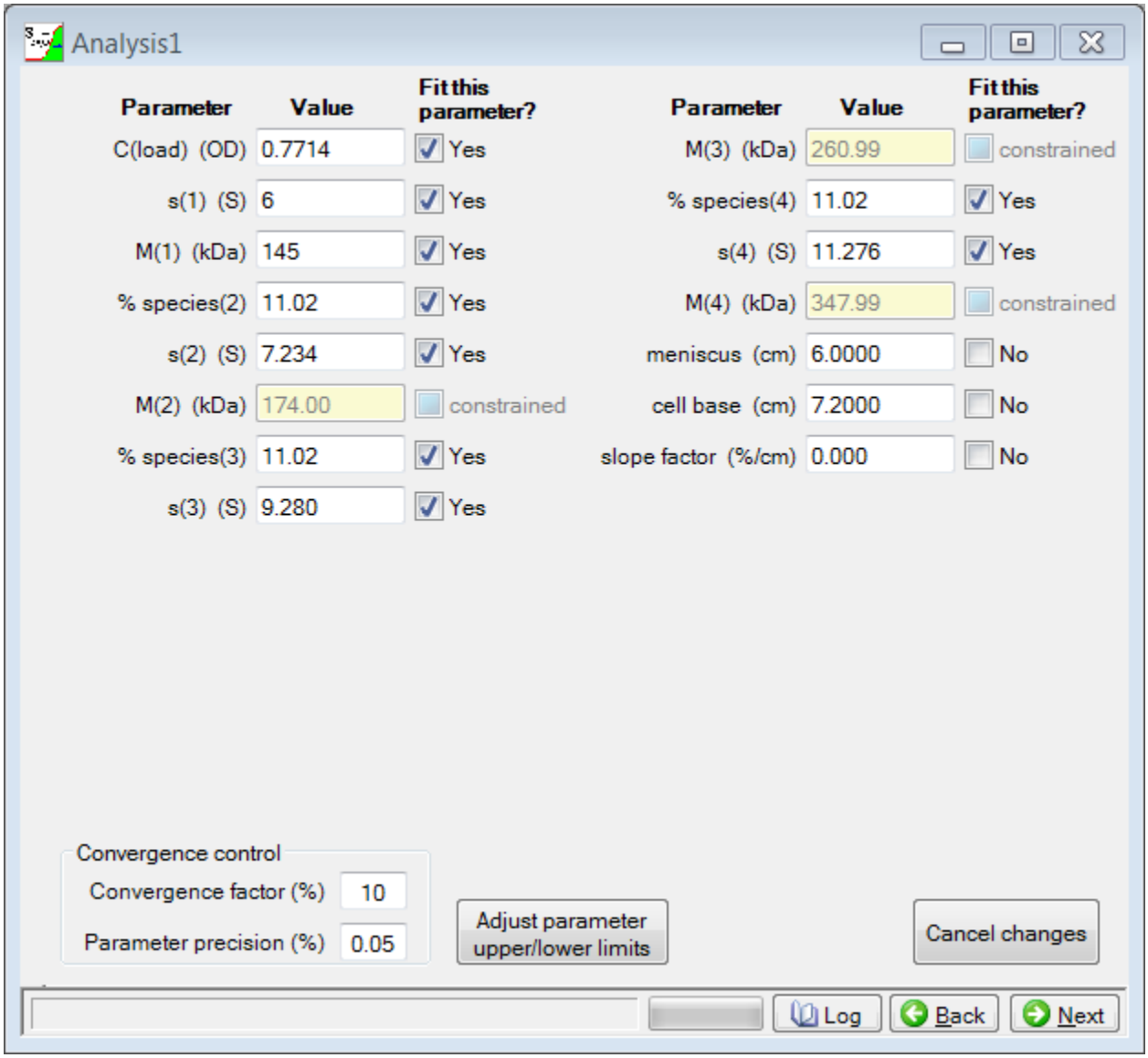

Set Molar Mass or Sedimentation Coefficient Constraints
Molar mass (s/D ratio) constraints
The radio buttons below set the species to which that molar mass (s/D) is linked
Ratio of molar masses*
Species 1 none 2 3 4 5 6 2.000
Species 2 none 1 3 4 5 6 2.000
Species 3 none 1 2 4 5 6 3.000
Species 4 none 1 2 3 5 6 4.000
Species 5 none 1 2 3 4 6 6.000
Species 6 none 1 2 3 4 5 8.000
* Enter -1 for constant f/f0 ratio mass scaling
Ok
Cancel changes
Clear mass constraints
Clear s constraints
Clear all constraints
Sedimentation coefficient constrain
The radio buttons below set the species to which that sedimentation coefficient is linked
Species 1 none 2 3 4 5 6
Species 2 none 1 3 4 5 6
Species 3 none 1 2 4 5 6
Species 4 none 1 2 3 5 6
Species 5 none 1 2 3 4 6
Species 6 none 1 2 3 4 5
Predicted sedimentation coefficient ratios
dimer 1.45
trimer - linear 1.75
trimer - triangular 1.86
tetramer - linear 2.00
tetramer - square planar 2.20
tetramer - tetrahedral 2.26

Potentially it might also make sense to constrain the sedimentation coefficient of the dimer, trimer, and tetramer species to ratios relative to the monomer that are appropriate for those oligomers, at least for an initial fit. That would automatically set the starting values of those sedimentation coefficients into a reasonable range, and those constraints could perhaps be released later in the fitting session. For this example we will not do this (but will manually input some reasonable guesses for those sedimentation coefficients). Therefore we are done with this form and can push the **Ok** button to enter our desired mass constraints and close this dialog.

Our fitting model is now complete so we can move forward to the **Set/alter fitting parameters page (Section 3.2.4)** and enter some initial guesses. This new set of data will trigger this dialog:



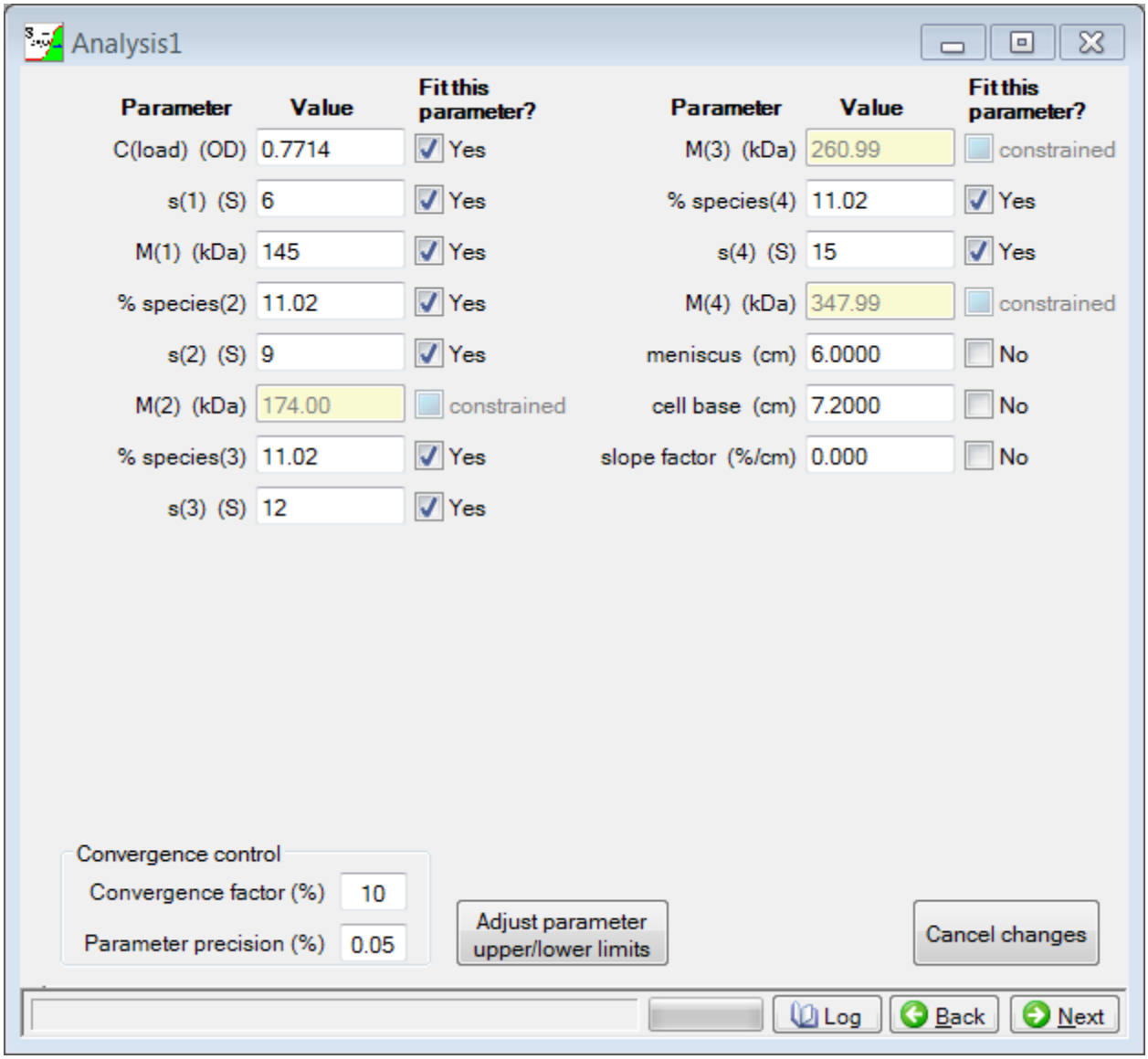
If we say yes the wizard will come up with these guesses (see the 'Easy case' section above for details about how these values are determined):



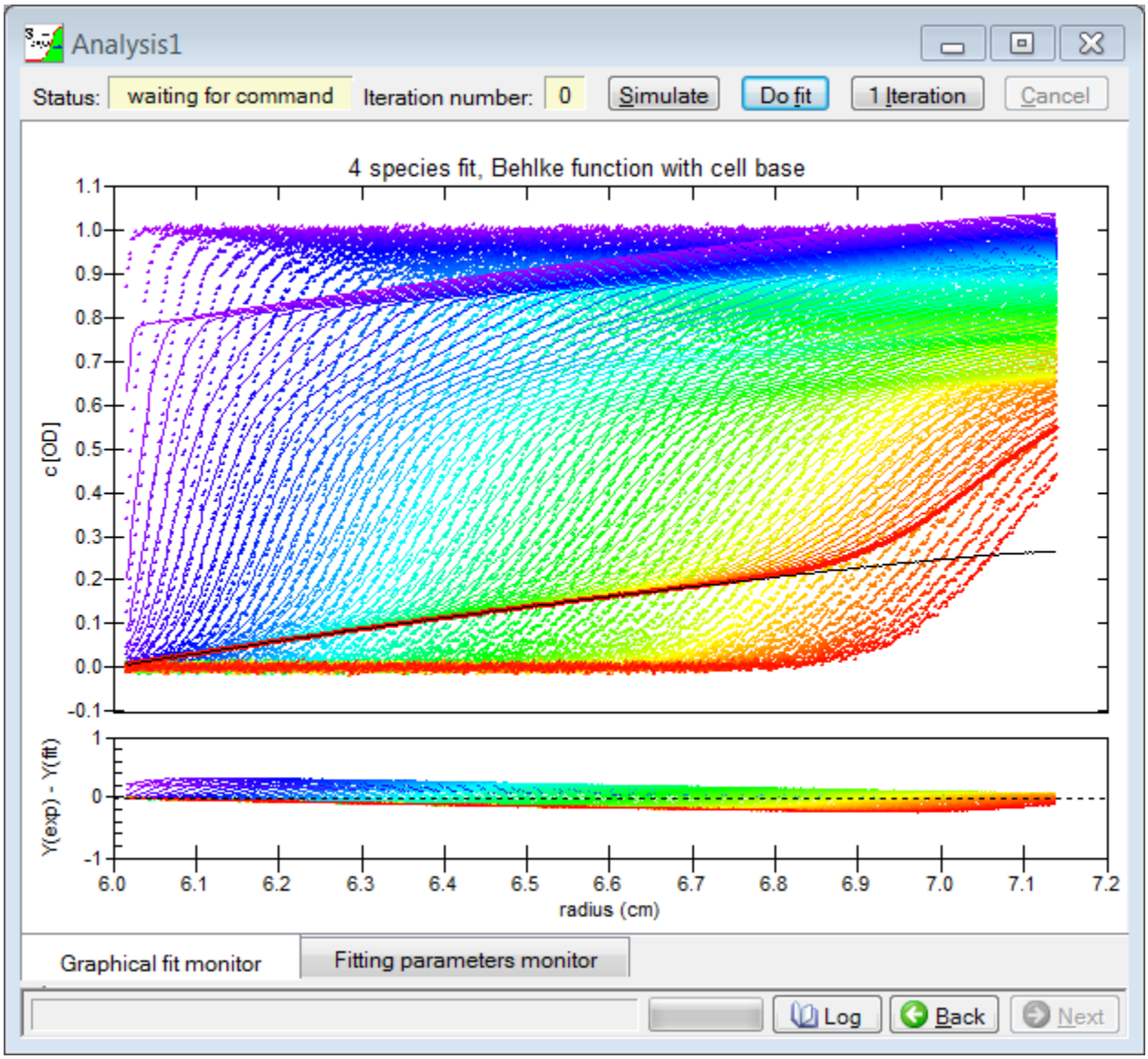
In this case the wizard has significantly under-estimated the monomer (species 1) sedimentation coefficient and mass. Changing those to a better estimate will, at the very least, significantly speed up the time to

convergence, and may actually be necessary for the fit to converge at all. Changing the initial guesses for the total concentration and species fractions to make them closer to the expected values would also speed up convergence, but usually is not required.

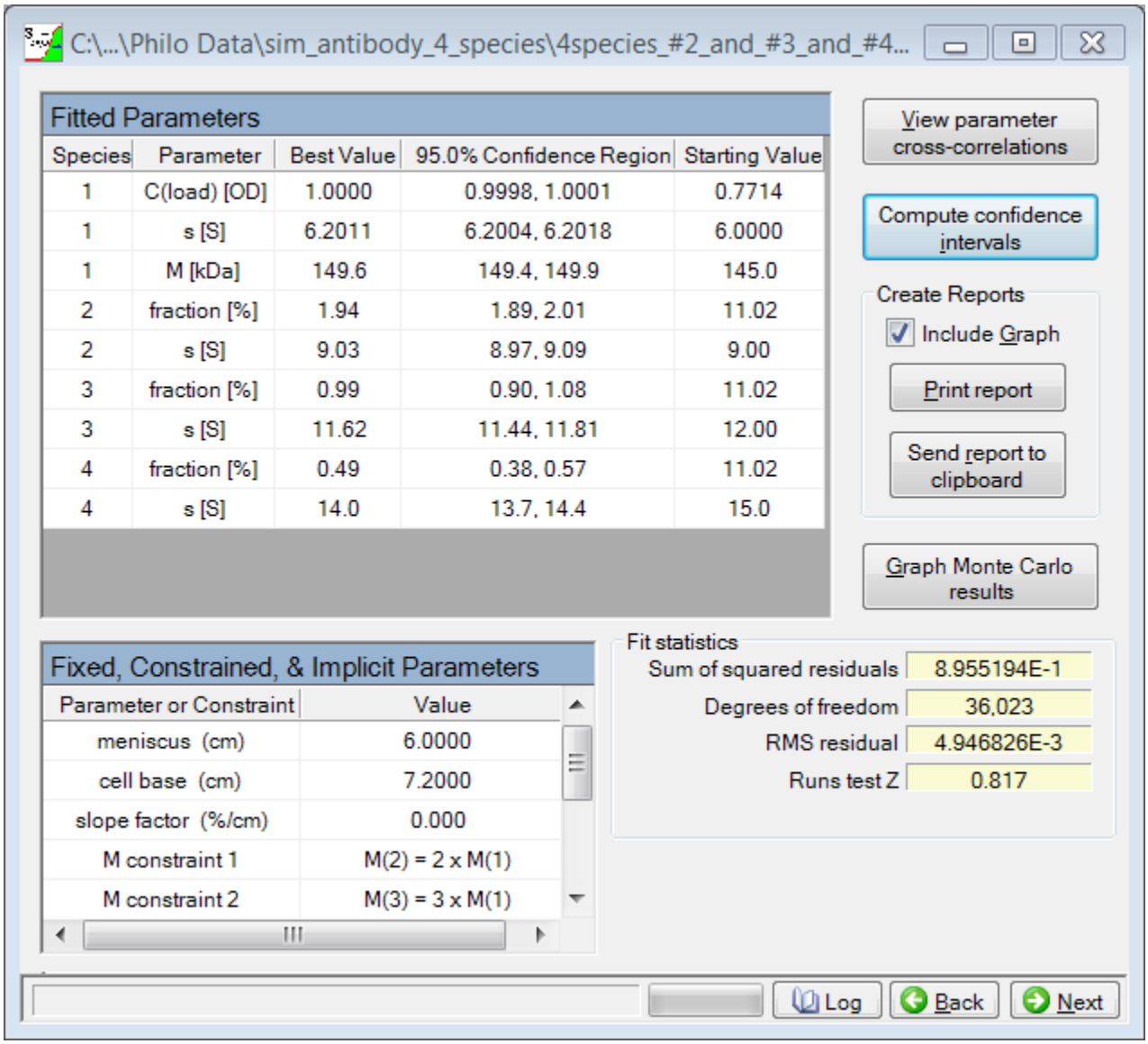
For our example we will change the sedimentation coefficient and mass of species 1 to initial guesses of 6.0 S and 145 kDa, and the sedimentation coefficients for species 2-4 to 9 S, 12 S, and 15 S. Here again for an initial fit of a complex mixture it is probably a good idea to hold the meniscus and cell base positions fixed (but they could be fitted later if desired). Thus our final page looks like this:



Moving forward to the **Perform least-squares fit page (Section 3.2.5)**, we see that those guesses don't provide a very good fit, but nonetheless we can proceed with the fit (the **Do fit** button).



The fit converges quickly (only 6 iterations, < 1 min). After evaluating the 95% confidence intervals using 200 rounds of Monte Carlo (taking ~25 min) we get the following final results:


C:\...\Philo Data\sim_antibody_4_species\4species_#2_and_#3_and_#4...
Fitted Parameters
Species | Parameter | Best Value | 95.0% Confidence Region | Starting Value
1 | C(load) [OD] | 1.0000 | 0.9998, 1.0001 | 0.7714
1 | s [S] | 6.2011 | 6.2004, 6.2018 | 6.0000
1 | M [kDa] | 149.6 | 149.4, 149.9 | 145.0
2 | fraction [%] | 1.94 | 1.89, 2.01 | 11.02
2 | s [S] | 9.03 | 8.97, 9.09 | 9.00
3 | fraction [%] | 0.99 | 0.90, 1.08 | 11.02
3 | s [S] | 11.62 | 11.44, 11.81 | 12.00
4 | fraction [%] | 0.49 | 0.38, 0.57 | 11.02
4 | s [S] | 14.0 | 13.7, 14.4 | 15.0
View parameter cross-correlations
Compute confidence intervals
Create Reports
Include Graph
Print report
Send report to clipboard
Graph Monte Carlo results
Fixed, Constrained, & Implicit Parameters
Parameter or Constraint | Value
meniscus (cm) | 6.0000
cell base (cm) | 7.2000
slope factor (%/cm) | 0.000
M constraint 1 | M(2) = 2 x M(1)
M constraint 2 | M(3) = 3 x M(1)
Fit statistics
Sum of squared residuals 8.955194E-1
Degrees of freedom 36,023
RMS residual 4.946826E-3
Runs test Z 0.817
Log
Back
Next

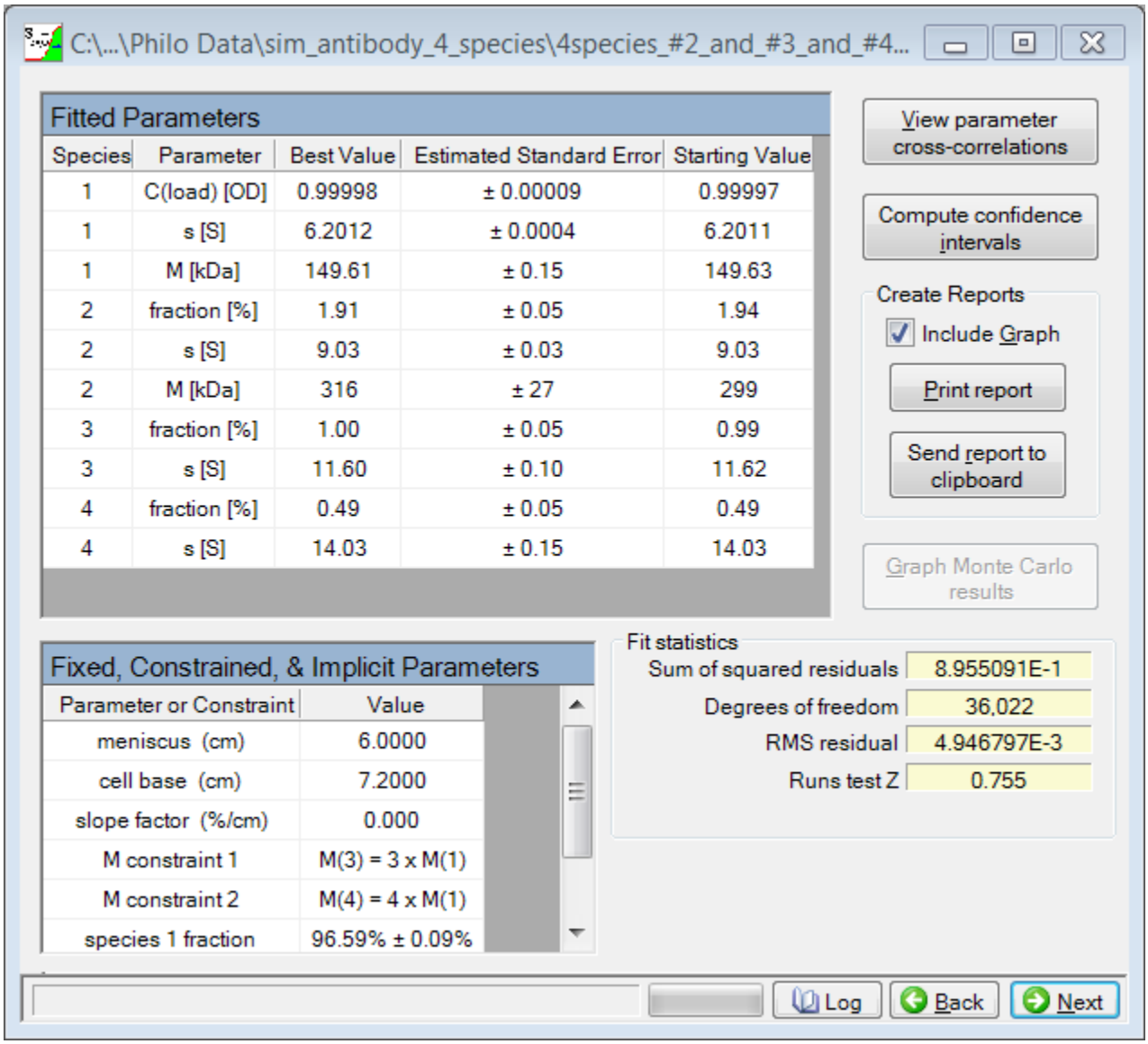
From this we see that the properties of all species are returned with good accuracy and reasonably good precision. But if this was a real sample, how would we know whether the assumptions about stoichiometry correct? One important point is that the best-fit sedimentation coefficients are hydrodynamically reasonable for those stoichiometries.

Can we remove the mass constraints?

Can we actually remove some or all of the mass constraints and get any useful information about the mass of minor species present at such low levels to confirm their stoichiometries? Since the dimer is present with the highest fraction, and precision of its fraction and sedimentation coefficient are higher than for the putative trimer and tetramer, it makes sense to try removing this constraint first.

To do that we navigate back to the **Select fitting model page (Section 3.2.3)**, push the **Set constraint values** button to bring up the constraints dialog, click the **None** radio button for the Species 2 link to remove that constraint, and then push **Ok**.

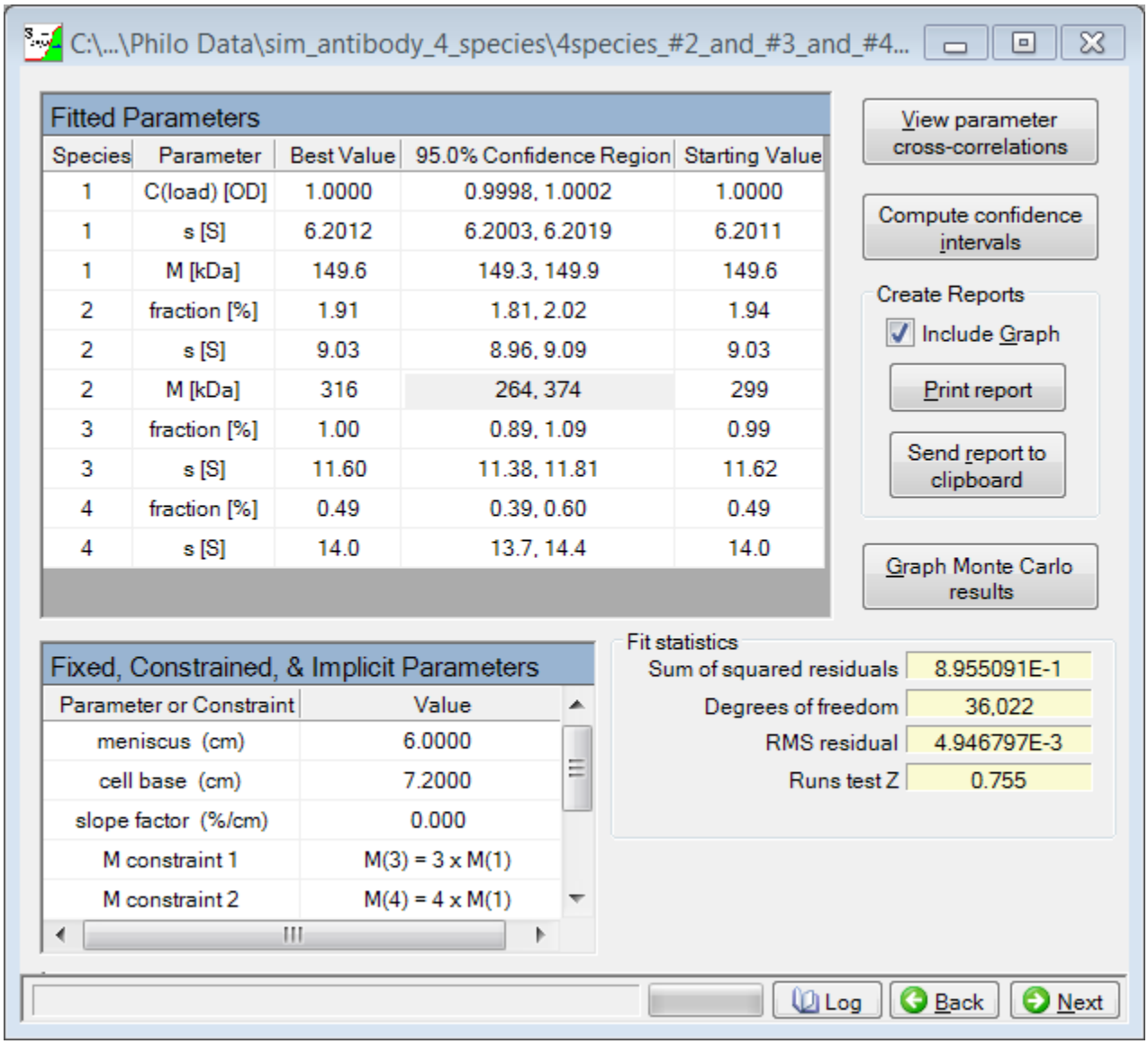
We can keep the previous fit results as our initial guesses, and therefore immediately re-fit without the constraint, giving this result:


C:\...\Philo Data\sim_antibody_4_species\4species_#2_and_#3_and_#4...
Fitted Parameters
Species | Parameter | Best Value | Estimated Standard Error | Starting Value
1 | C(load) [OD] | 0.99998 | ± 0.00009 | 0.99997
1 | s [S] | 6.2012 | ± 0.0004 | 6.2011
1 | M [kDa] | 149.61 | ± 0.15 | 149.63
2 | fraction [%] | 1.91 | ± 0.05 | 1.94
2 | s [S] | 9.03 | ± 0.03 | 9.03
2 | M [kDa] | 316 | ± 27 | 299
3 | fraction [%] | 1.00 | ± 0.05 | 0.99
3 | s [S] | 11.60 | ± 0.10 | 11.62
4 | fraction [%] | 0.49 | ± 0.05 | 0.49
4 | s [S] | 14.03 | ± 0.15 | 14.03
View parameter cross-correlations
Compute confidence intervals
Create Reports
Include Graph
Print report
Send report to clipboard
Graph Monte Carlo results
Fixed, Constrained, & Implicit Parameters
Parameter or Constraint | Value
meniscus (cm) | 6.0000
cell base (cm) | 7.2000
slope factor (%/cm) | 0.000
M constraint 1 | M(3) = 3 x M(1)
M constraint 2 | M(4) = 4 x M(1)
species 1 fraction | 96.59% ± 0.09%
Fit statistics
Sum of squared residuals | 8.955091E-1
Degrees of freedom | 36,022
RMS residual | 4.946797E-3
Runs test Z | 0.755
Log
Back
Next

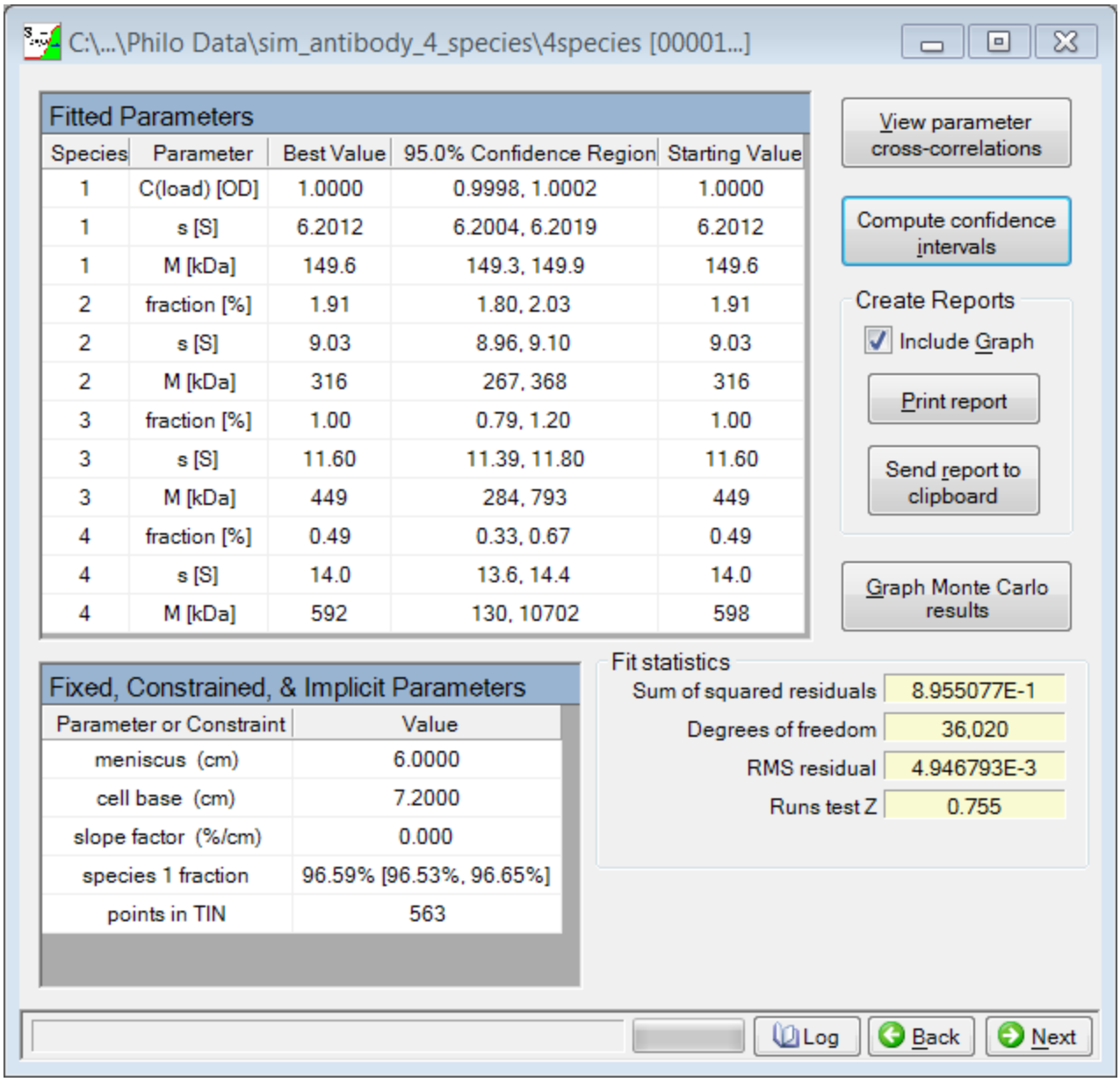
We see that the estimated standard error for the dimer molar mass (±27 kDa) does suggest that the precision is high enough to confirm this species is a dimer.

However the results of the Monte Carlo analysis below give a 95% confidence interval of 264 to 374 kDa (1.76 to 2.49 monomer units) for species 2, and thus a stoichiometry of trimer is only marginally excluded at a 95% confidence level.

Note also that removing the constraint results in poorer accuracy and precision for the fraction of species 2 (and slightly lower precision for species 3 and 4 as well).


C:\...\Philo Data\sim_antibody_4_species\4species_#2_and_#3_and_#4...
Fitted Parameters
Species | Parameter | Best Value | 95.0% Confidence Region | Starting Value
1 | C(load) [OD] | 1.0000 | 0.9998, 1.0002 | 1.0000
1 | s [S] | 6.2012 | 6.2003, 6.2019 | 6.2011
1 | M [kDa] | 149.6 | 149.3, 149.9 | 149.6
2 | fraction [%] | 1.91 | 1.81, 2.02 | 1.94
2 | s [S] | 9.03 | 8.96, 9.09 | 9.03
2 | M [kDa] | 316 | 264, 374 | 299
3 | fraction [%] | 1.00 | 0.89, 1.09 | 0.99
3 | s [S] | 11.60 | 11.38, 11.81 | 11.62
4 | fraction [%] | 0.49 | 0.39, 0.60 | 0.49
4 | s [S] | 14.0 | 13.7, 14.4 | 14.0
View parameter cross-correlations
Compute confidence intervals
Create Reports
Include Graph
Print report
Send report to clipboard
Graph Monte Carlo results
Fixed, Constrained, & Implicit Parameters
Parameter or Constraint | Value
meniscus (cm) | 6.0000
cell base (cm) | 7.2000
slope factor (%/cm) | 0.000
M constraint 1 | M(3) = 3 x M(1)
M constraint 2 | M(4) = 4 x M(1)
Fit statistics
Sum of squared residuals 8.955091E-1
Degrees of freedom 36,022
RMS residual 4.946797E-3
Runs test Z 0.755
Log
Back
Next

It turns out that is possible to remove all the constraints and still get the fit to converge (when beginning with the good starting values from the constrained fit). Those results with 95% confidence limits are shown below:


C:\...\Philo Data\sim_antibody_4_species\4species [00001...]
Fitted Parameters
Species Parameter Best Value 95.0% Confidence Region Starting Value
1 C(load) [OD] 1.0000 0.9998, 1.0002 1.0000
1 s [S] 6.2012 6.2004, 6.2019 6.2012
1 M [kDa] 149.6 149.3, 149.9 149.6
2 fraction [%] 1.91 1.80, 2.03 1.91
2 s [S] 9.03 8.96, 9.10 9.03
2 M [kDa] 316 267, 368 316
3 fraction [%] 1.00 0.79, 1.20 1.00
3 s [S] 11.60 11.39, 11.80 11.60
3 M [kDa] 449 284, 793 449
4 fraction [%] 0.49 0.33, 0.67 0.49
4 s [S] 14.0 13.6, 14.4 14.0
4 M [kDa] 592 130, 10702 598
View parameter cross-correlations
Compute confidence intervals
Create Reports
Include Graph
Print report
Send report to clipboard
Graph Monte Carlo results
Fixed, Constrained, & Implicit Parameters
Parameter or Constraint Value
meniscus (cm) 6.0000
cell base (cm) 7.2000
slope factor (%/cm) 0.000
species 1 fraction 96.59% [96.53%, 96.65%]
points in TIN 563
Fit statistics
Sum of squared residuals 8.955077E-1
Degrees of freedom 36,020
RMS residual 4.946793E-3
Runs test Z 0.755
Log
Back
Next

Not too surprisingly, this result shows that the molar masses for species 3 and 4 are quite poorly defined. At 95% confidence species 3 could be anything from dimer to pentamer, and species 4 could be anything from monomer to 71-mer.

Further, removing those constraints again lowers the accuracy and precision of the species fractions and sedimentation coefficients. Thus trying to obtain too much information from the fit not only fails to get useful information about those molar masses, but also compromises all the other information from the fit. This is exactly why constraints can be quite useful!

A final point is that although the masses of species 3 and 4 cannot be well defined from their boundary widths, one could set some tighter "reality check" limits based on their sedimentation coefficients and simple hydrodynamic principles. For example, although the fit confidence interval allows species 4 to be a monomer, a 150 kDa monomer cannot possibly sediment as fast as 13.6 S (the lower confidence limit for its sedimentation coefficient). Similarly if we limit its f/f_0 ratio to a maximum of 3 then the upper bound of 14.4 S implies species 4 cannot be larger than 9-mer.

2.3 Tutorial: Using a c(s) distribution with SVEDBERG

noSVEDBERG has several features which make it easy to translate a *c(s)* distribution calculated in SEDFIT into a discrete species model. This tutorial will illustrate how this is done.

Why would I want to do this?

The great strength of the *c(s)* method is that it requires no *a priori* assumptions about how many species are present in a sample. However it also has some significant drawbacks:

1. It never provides confidence limits for the total concentration, species fractions, or sedimentation coefficients of individual species.
2. It always assumes that all species have the same f/f_0 ratio. This is likely not true for different oligomers of a single protein, for mixtures of two or more proteins, or when a sample contains impurities or protein fragments. Mass estimates for individual peaks (from the CTRL-M command or translation into a *c(M)* distribution) will not be correct unless this assumption is strictly obeyed.
3. The *c(s)* method implicitly assumes that all species in a mixture have the same partial specific volume.

When the *c(s)* distribution indicates that a fairly small number of species are present, and when it is reasonable to treat the sample as an ideal non-interacting mixture, then it becomes possible to re-fit the data using a discrete species model. Doing this allows you to:

1. Obtain confidence limits for the total concentration and species fractions (or for the concentration of individual species) and for the sedimentation coefficients
2. Try removing or altering the constant f/f_0 ratio constraints. For example, it may be appropriate to constrain the mass ratios between species to values appropriate for different oligomers. For species which represent a reasonably high fraction of the total it will likely be possible to remove all constraints and directly fit for the mass of that species (and to evaluate the precision to which that mass can be determined).
3. For mixtures with more than one protein, if you can assign the appropriate partial specific volume to each species (based on its sedimentation coefficient) then you can get the correct molar mass for individual components.

How does the translation to a discrete species model work?

The basic idea is that each peak in the *c(s)* distribution becomes one discrete species. Like in the *c(s*)* model, when there are 2 or more species the masses of the species are all linked together by a common f/f_0 ratio. However this is implemented by designating one species as the major or 'controlling' species, with the unknown molar mass of that species as a fitting parameter, and then using **a f/f0 ratio constraint (Section 3.2.3.7)** to determine the masses of the remaining species from their sedimentation coefficients. That is, as the mass and sedimentation coefficient of the 'controlling' species is varied to find the best fit, its current f/f_0 ratio is computed and then applied to the other species to specify their masses for that iteration of the fit.

The 'controlling' species should be the one whose hydrodynamic shape has the most influence on the quality of the fit, and therefore it usually should be the one with the largest peak fraction (highest concentration).

Note that the translated discrete species model will not be logically exactly equivalent to the *c(s)* fit:

- One difference is that peaks in the *c(s)* distribution are normally wider than a single data point, and thus one 'logical peak' in the *c(s)* distribution really was represented as a group of species with a range of sedimentation coefficients and masses, but that group becomes one average species in SVEDBERG.
- Another important difference is that the *c(s)* fit is using a pre-determined set of possible sedimentation coefficients, and logically the fitter is only determining what concentration is present at each of those sedimentation coefficient values. In contrast, in the discrete species model the fitter is trying to choose among an infinite number of possible values for the sedimentation coefficient for each species. This is a subtle distinction, but one that is quite important for species which are present at very low levels and

hence whose sedimentation coefficient will have low precision. Largely due to this difference, in some cases it will not be possible for the discrete species to converge on the properties of a low-level *c(s)* peak; a choice between fixed sedimentation coefficient values of 1.0, 1.1., or 1.2 S, for example, is much easier than trying to converge on a continuously-variable best fit value within that range.

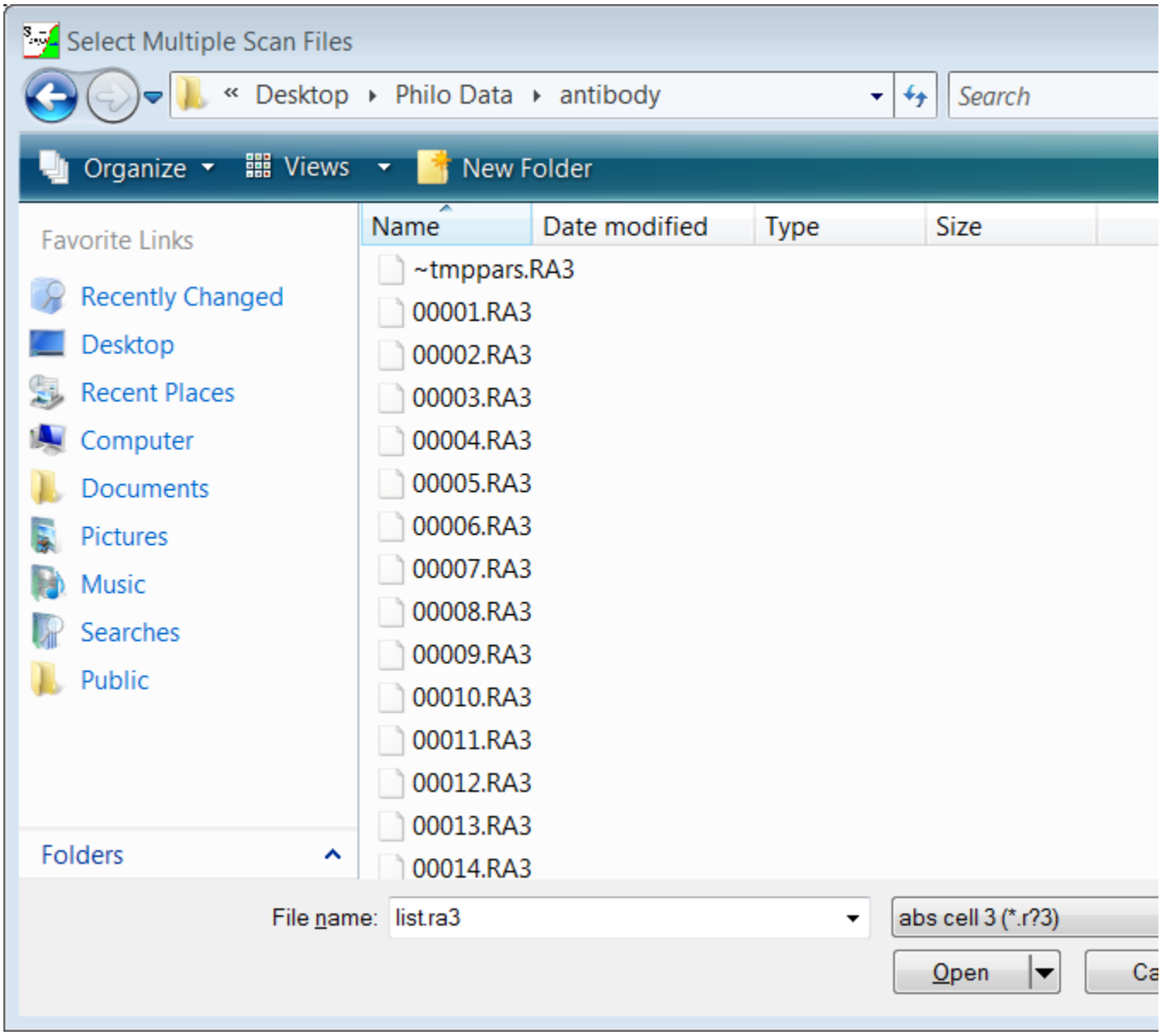
What procedure do I follow to load the scans and import the *c(s)* distribution?

When you do the *c(s)* fit you must save the *c(s)* distribution as a text file using the SEDFIT **Data...Save Continuous Distribution** command (these are normally saved with a .dat file extension).

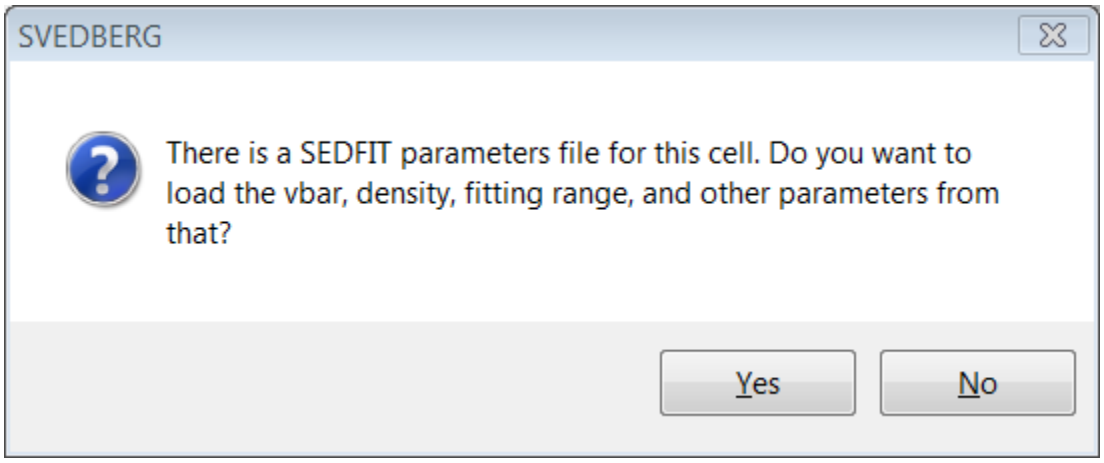
Then in SVEDBERG:

(1) Open a new analysis document (**File...New Analysis Window** or the equivalent button on the **Toolbar (Section 3.1.2)**).

(2) Push the **Load new scan files button (Section 3.2.1.2)** and then type in the filename name of the SEDFIT list file generated by the *c(s)* analysis. In this example it is 'list.ra3' since we will be using absorbance data from cell 3, but it would be 'list.ip2' for interference data from cell 2, *etc*. Using the SEDFIT list file will automatically load the same scan files that were used for the *c(s)* analysis (assuming you did not analyze different scans from that experiment using SEDFIT at some later time.)

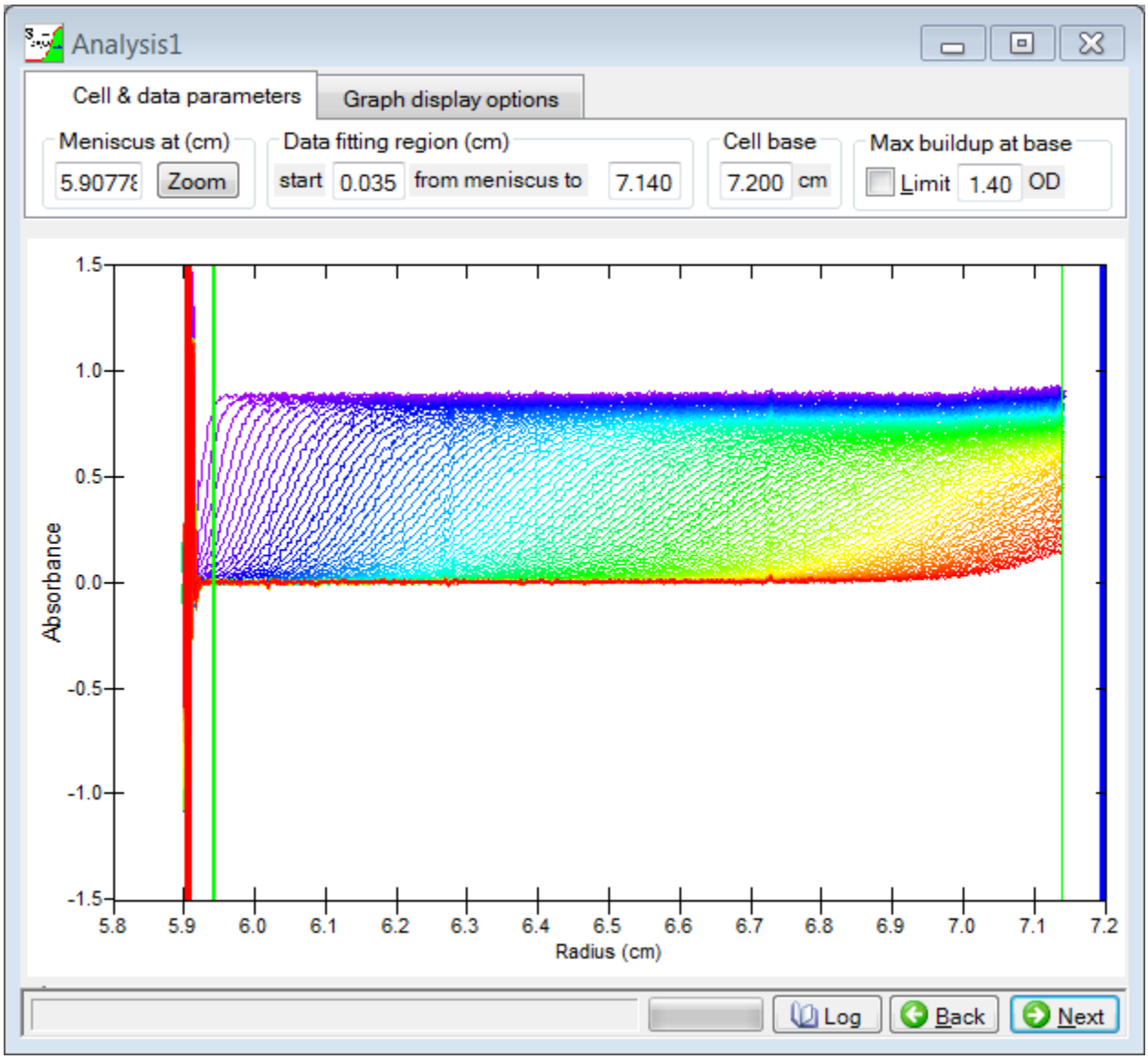


(3) After the scans load you will see the following dialog:

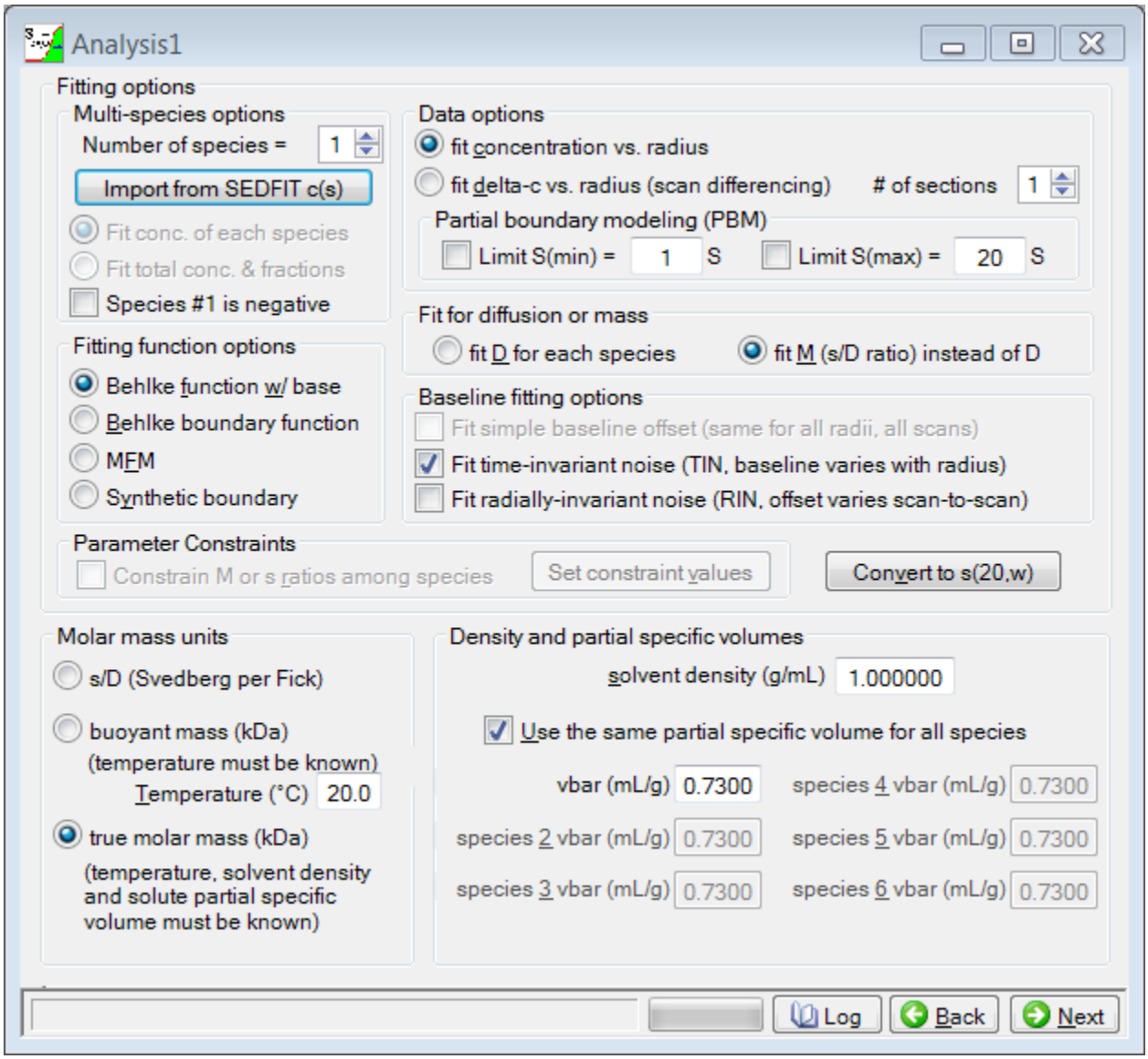

SVEDBERG
There is a SEDFIT parameters file for this cell. Do you want to load the vbar, density, fitting range, and other parameters from that?
Yes
No

Answer 'Yes' and the meniscus and cell base positions, radial range of data used in the fitting, buffer density and viscosity, and partial specific volume values will be set to the same values used for the *c(s)* analysis. (These values are read in from the SEDFIT parameter file named '~tmppars.ip3' in this case.)

(4) When you push the **Next button (C:\Users\Public\Documents\Visual Studio Projects\Svedberg Help Source\{46451615-072B-4FCC-B31F-82953CB24396}#3)** to move to the **(C:\Users\Public\Documents\Visual Studio Projects\Svedberg Help Source\{14FCF551-188C-4C48-9E48-0B1B6957B1E6}#3)Set meniscus & data region page (C:\Users\Public\Documents\Visual Studio Projects\Svedberg Help Source\{14FCF551-188C-4C48-9E48-0B1B6957B1E6}#3)** you will see that the meniscus, cell base, and fitting region have already been set.



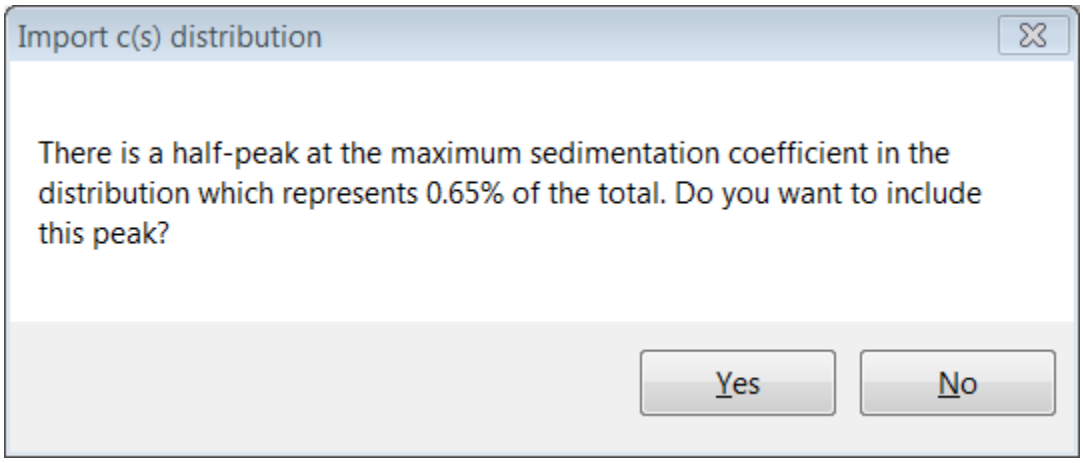
(5) Move forward to the **Select fitting model page (Section 3.2.3)**, then push the **Import from SEDFIT c(s) button (Section 3.2.3.2)**.


Analysis1
Fitting options
Multi-species options
Number of species = 1
Import from SEDFIT c(s)
Fit conc. of each species
Fit total conc. & fractions
Species #1 is negative
Fitting function options
Behlke function w/ base
Behlke boundary function
MEM
Synthetic boundary
Data options
fit concentration vs. radius
fit delta-c vs. radius (scan differencing)
# of sections 1
Partial boundary modeling (PBM)
Limit S(min) = 1 S
Limit S(max) = 20 S
Fit for diffusion or mass
fit D for each species
fit M (s/D ratio) instead of D
Baseline fitting options
Fit simple baseline offset (same for all radii, all scans)
Fit time-invariant noise (TIN, baseline varies with radius)
Fit radially-invariant noise (RIN, offset varies scan-to-scan)
Parameter Constraints
Constrain M or s ratios among species
Set constraint values
Convert to s(20,w)
Molar mass units
s/D (Svedberg per Fick)
buoyant mass (kDa)
(temperature must be known)
Temperature (°C) 20.0
true molar mass (kDa)
(temperature, solvent density and solute partial specific volume must be known)
Density and partial specific volumes
solvent density (g/mL) 1.000000
Use the same partial specific volume for all species
vbar (mL/g) 0.7300
species 4 vbar (mL/g) 0.7300
species 2 vbar (mL/g) 0.7300
species 5 vbar (mL/g) 0.7300
species 3 vbar (mL/g) 0.7300
species 6 vbar (mL/g) 0.7300
Log
Back
Next

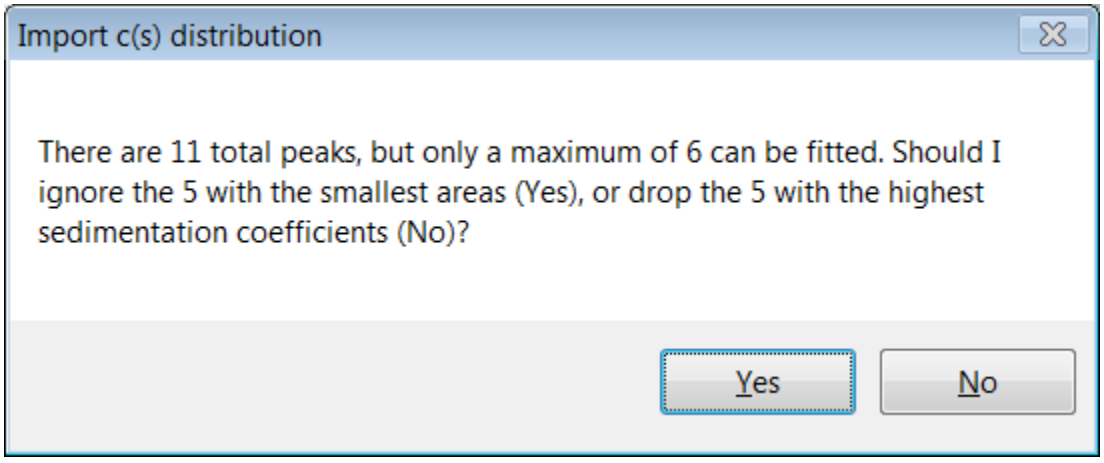
This will bring up a **File Open** dialog for you to locate the file where the *c(s)* distribution was stored (again by default the dialog expects this file to have a .dat extension, and to be located in the same folder from which the scan files were loaded).

(6) After the *c(s)* distribution is loaded a program wizard locates the peak boundaries then integrates the peaks. Depending on those results a series of several dialog boxes may be displayed.

The program first checks whether half-peaks are present at the lower or upper end of the sedimentation coefficient range (that is, whether the distribution is still increasing as the limit is reached). If so, those *might* be artifacts you don't want to include in the fit, so you will see one or two dialog boxes similar to this one:

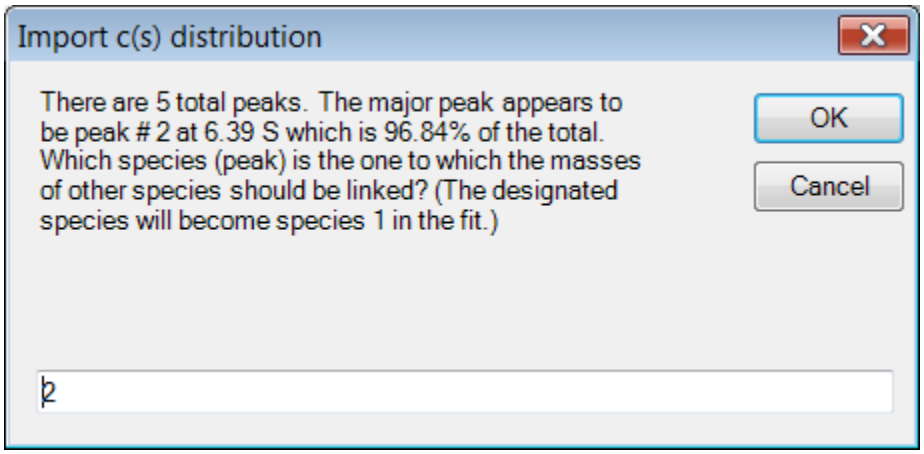

Import c(s) distribution
There is a half-peak at the maximum sedimentation coefficient in the distribution which represents 0.65% of the total. Do you want to include this peak?
Yes
No

Next the program checks whether the total number of peaks (less any half-peaks which were removed) exceeds six, the maximum number of discrete species SVEDBERG will fit. If so, you will see a dialog similar to this one:


Import c(s) distribution
There are 11 total peaks, but only a maximum of 6 can be fitted. Should I ignore the 5 with the smallest areas (Yes), or drop the 5 with the highest sedimentation coefficients (No)?
Yes
No

You may choose one of two options for eliminating the excess peaks, but in this situation it is likely this discrete species fit will not work well (or not be meaningful). Sometimes it may be appropriate to group together two or more minor species (for example trace aggregates sedimenting faster than the species of main interest) into a single, average species.

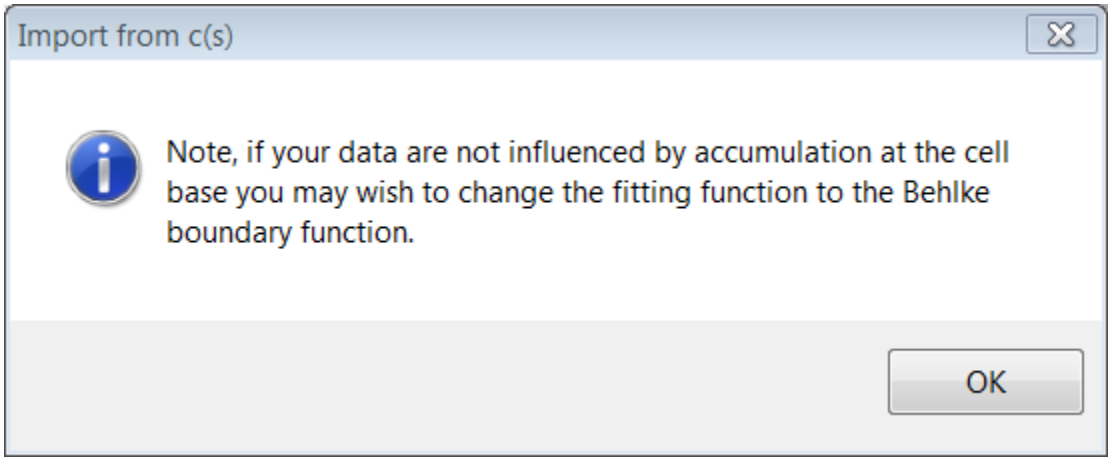
If at this point the c(s) distribution contains more than one peak you will then see the following dialog:


Import c(s) distribution
There are 5 total peaks. The major peak appears to be peak # 2 at 6.39 S which is 96.84% of the total. Which species (peak) is the one to which the masses of other species should be linked? (The designated species will become species 1 in the fit.)
OK
Cancel
2

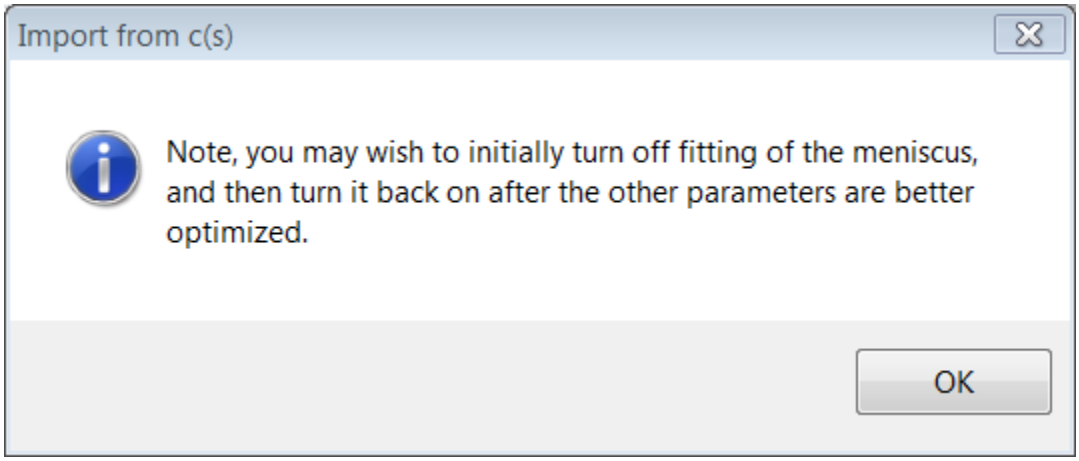
The purpose of this dialog is to designate which species will be the one whose molar mass will be directly fitted (the 'controlling' species discussed earlier), with masses for the others linked to that one through the constant

f/f_0 ratio constraints. Usually this should be the species present at highest concentration, and the number of that peak will be automatically entered in this dialog box as the default.

If your SEDFIT fit was not fitting the position of the cell base, that suggests that buildup at the cell base may not be important for this sample, and therefore you will see the following reminder:

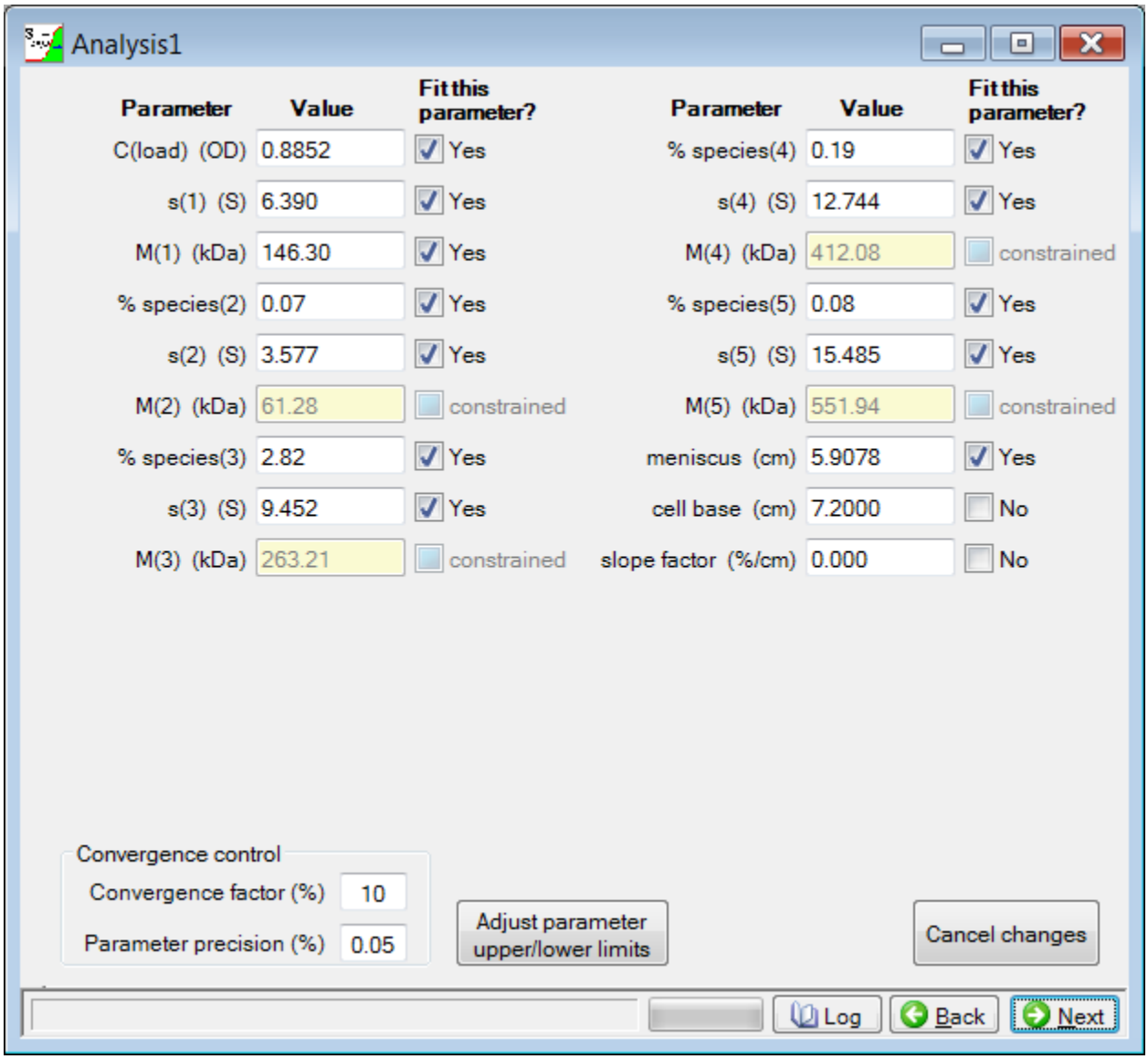

Import from c(s)
Note, if your data are not influenced by accumulation at the cell base you may wish to change the fitting function to the Behlke boundary function.
OK

Finally, if the SEDFIT analysis included fitting the meniscus position, you may see a dialog containing this reminder (sometimes combined with the preceding one):

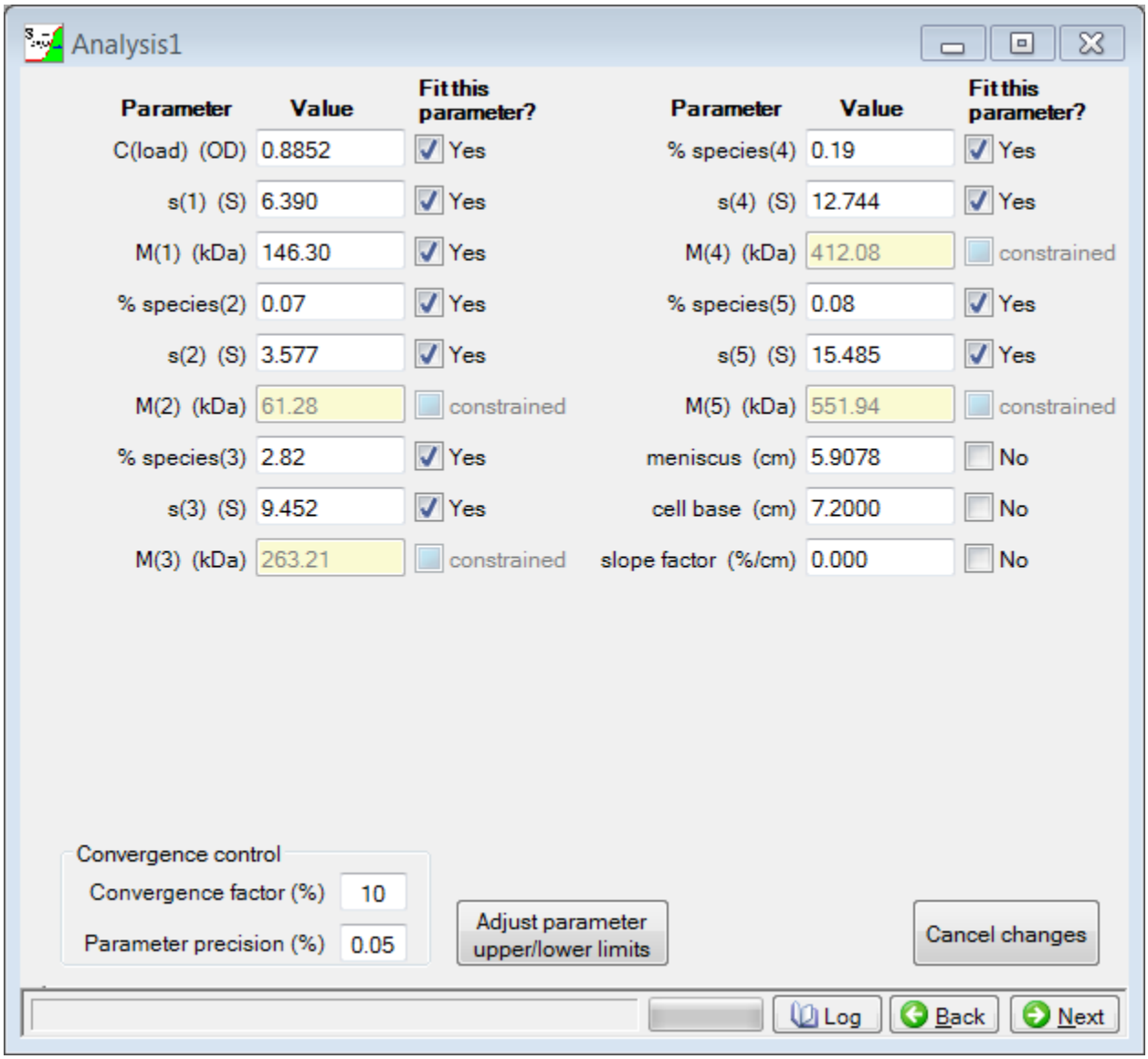

Import from c(s)
Note, you may wish to initially turn off fitting of the meniscus, and then turn it back on after the other parameters are better optimized.
OK

Because the results from integrating the *c(s)* peaks will not necessarily give good initial guesses for the SVEDBERG discrete species fit, when some of the species are present in very low amounts this suggestion to temporarily turn off fitting of the meniscus can help the SVEDBERG fit to converge (by helping it to avoid reducing a trace species to zero concentration as it hunts for an optimum solution).

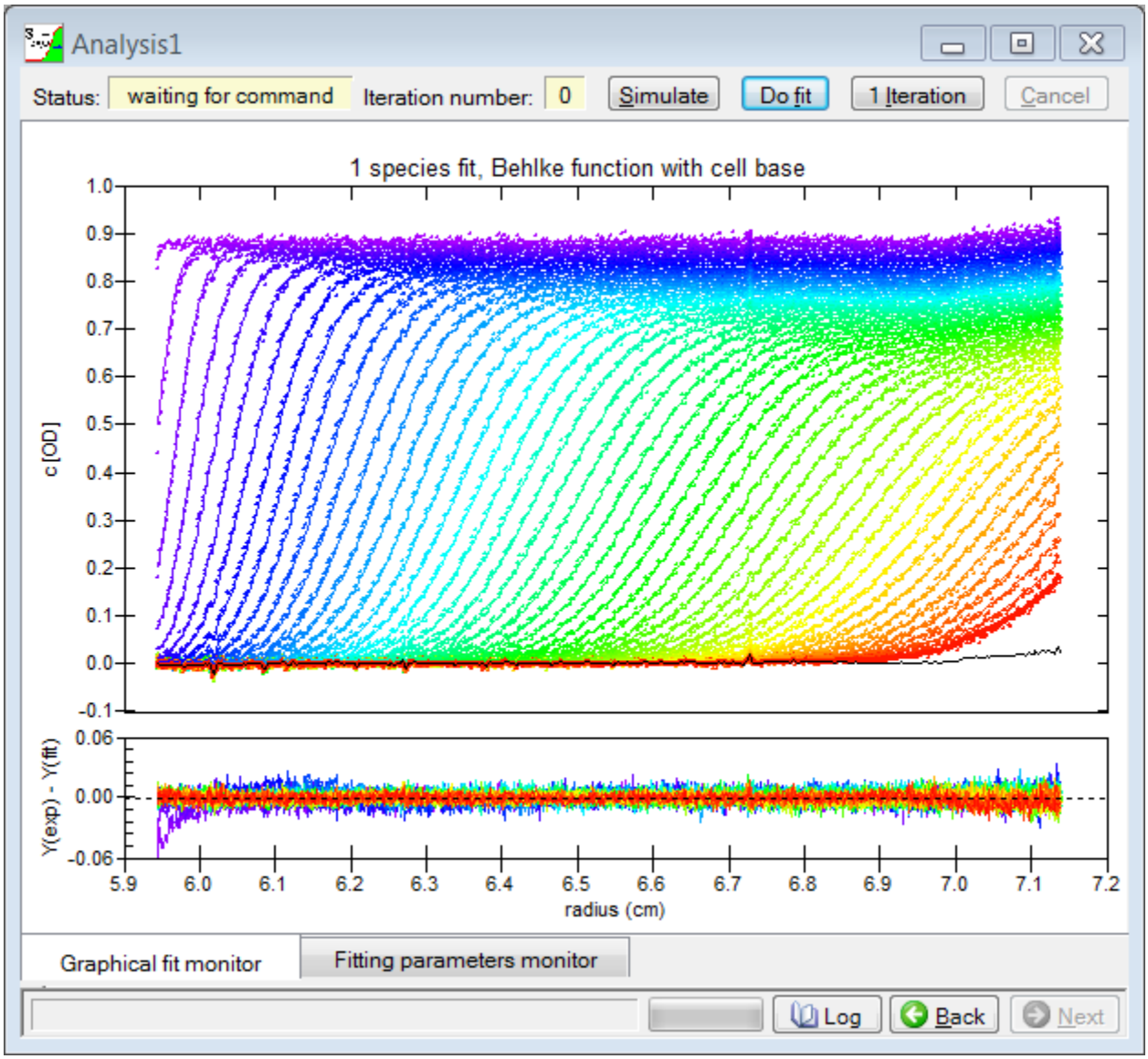
(7) Move forward to the **Set/alter fitting parameters page (Section 3.2.4)**.



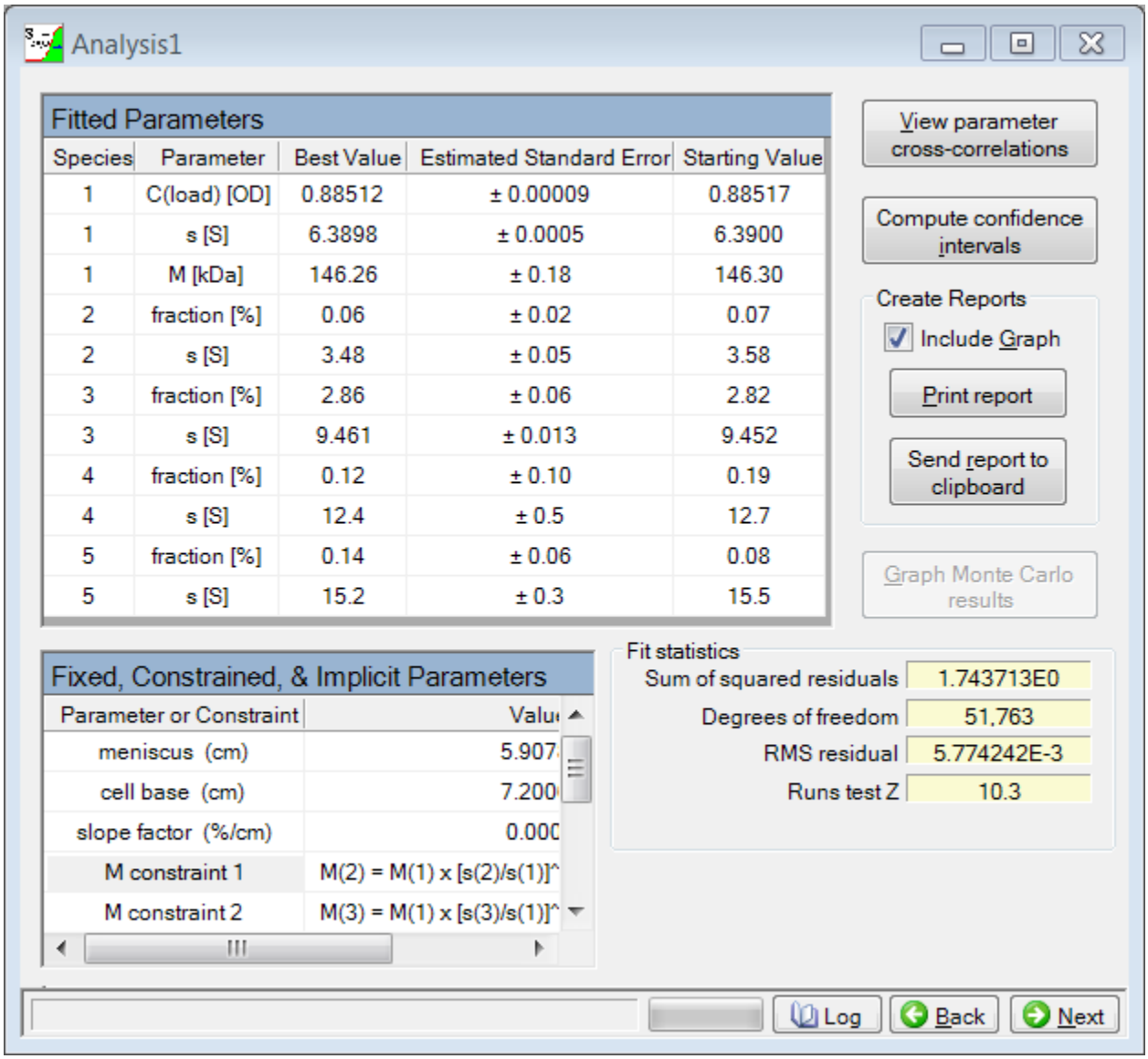
Note that for this sample the fractions for species 2, 4 and 5 are very low (0.07%, 0.19%, and 0.08%). This is exactly the situation where there is a danger that the initial guesses are not close enough and the first iteration or two will push the fraction of one of these species so close to zero that its sedimentation coefficient becomes undeterminable and the fit will not converge. This would not actually happen for this example, but nonetheless to illustrate the strategy we will temporarily turn off fitting of the meniscus, which helps avoid this problem. Clicking the check box next to the meniscus parameter will hold the current value fixed, making this page look like this:



(8) Now move forward to the **Perform least-squares fit page (Section 3.2.5)** and push the **Do fit button (Section 3.2.5.1)** to start the fit.

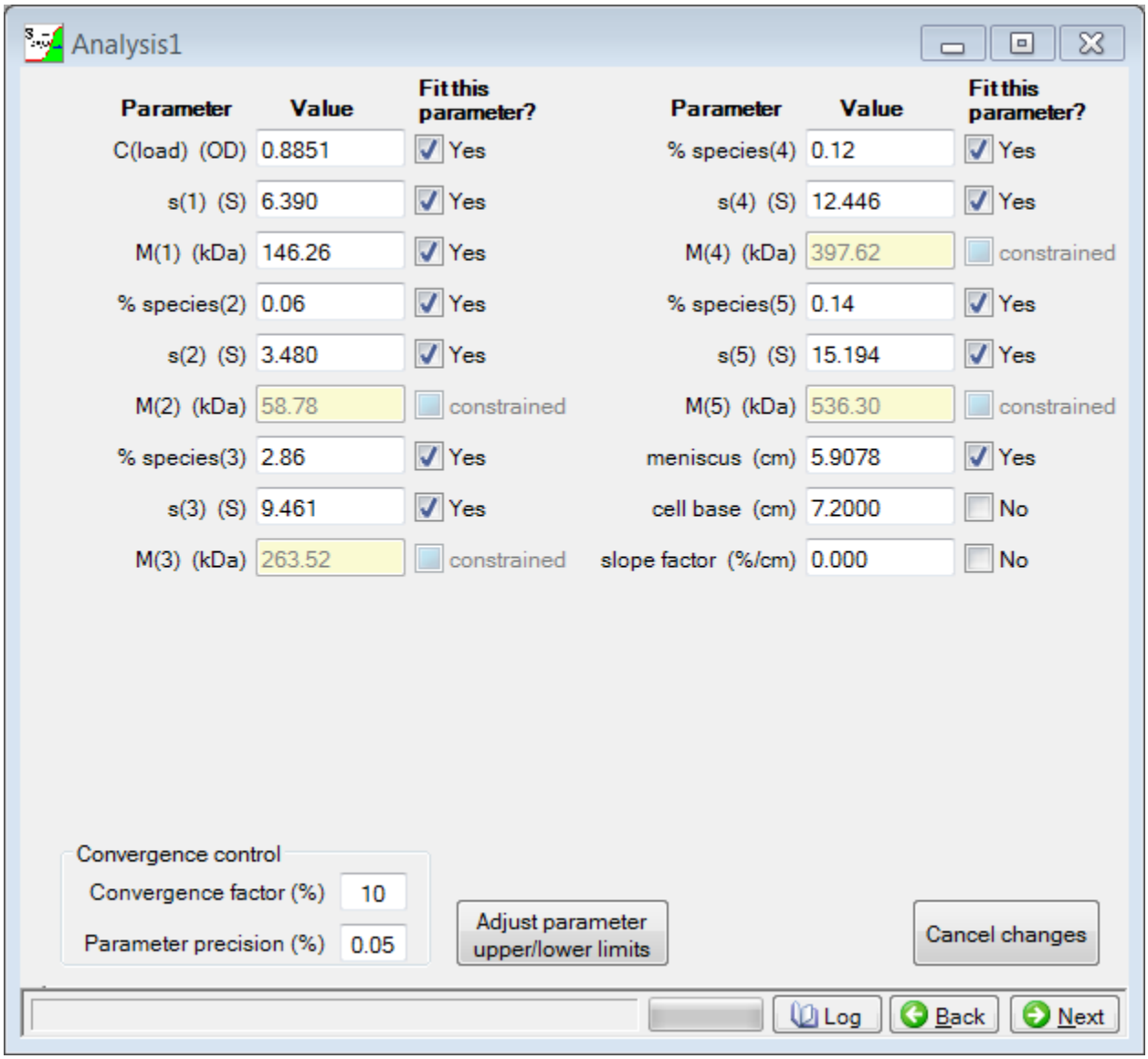


When the fit converges (after only 4 iterations in this case) the **Report fit results page (Section 3.2.6)** looks like this:

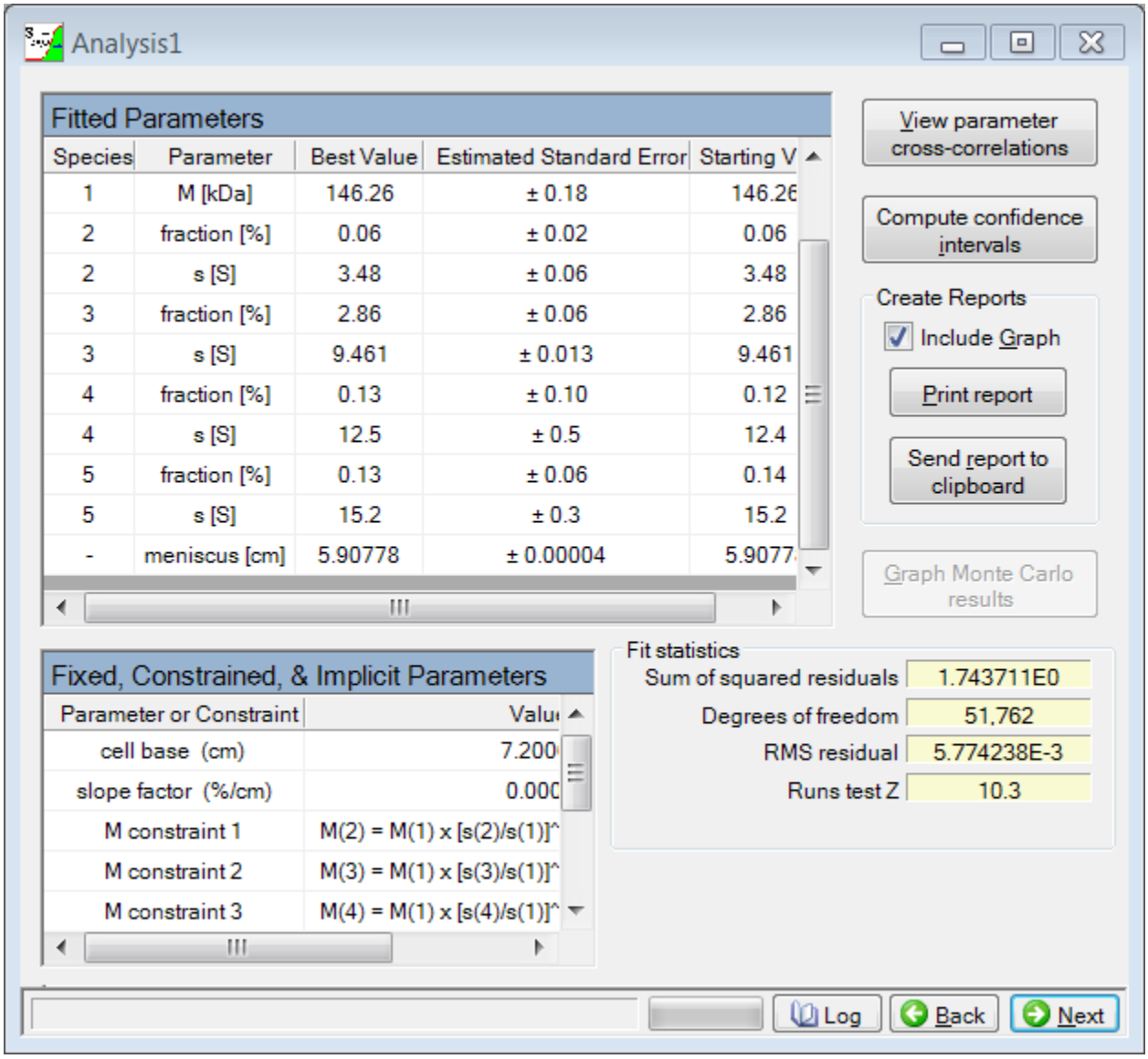

Analysis1
Fitted Parameters
Species Parameter Best Value Estimated Standard Error Starting Value
1 C(load) [OD] 0.88512 ± 0.00009 0.88517
1 s [S] 6.3898 ± 0.0005 6.3900
1 M [kDa] 146.26 ± 0.18 146.30
2 fraction [%] 0.06 ± 0.02 0.07
2 s [S] 3.48 ± 0.05 3.58
3 fraction [%] 2.86 ± 0.06 2.82
3 s [S] 9.461 ± 0.013 9.452
4 fraction [%] 0.12 ± 0.10 0.19
4 s [S] 12.4 ± 0.5 12.7
5 fraction [%] 0.14 ± 0.06 0.08
5 s [S] 15.2 ± 0.3 15.5
View parameter cross-correlations
Compute confidence intervals
Create Reports
Include Graph
Print report
Send report to clipboard
Graph Monte Carlo results
Fixed, Constrained, & Implicit Parameters
Parameter or Constraint
meniscus (cm)
cell base (cm)
slope factor (%/cm)
M constraint 1
M(2) = M(1) x [s(2)/s(1)]^
M constraint 2
M(3) = M(1) x [s(3)/s(1)]^
Fit statistics
Sum of squared residuals 1.743713E0
Degrees of freedom 51,763
RMS residual 5.774242E-3
Runs test Z 10.3
Log
Back
Next

Note that the constant f/f_0 ratio constraints used to link the molar mass of species 2-5 to the (fitted) molar mass of species 1 (antibody monomer in this case) are noted in the **Fixed, constrained, & implicit parameters table (Section 3.2.6.2)**.

(9) The above fit is correct, but remember that to help it converge we temporarily turned off fitting of the meniscus position. Thus to make our fit better parallel the *c(s)* fit we should turn that back on and re-fit. This can be done by pushing the **Back button (C:\Users\Public\Documents\Visual Studio Projects\Svedberg Help Source\{46451615-072B-4FCC-B31F-82953CB24396}#3)** twice to return to the **Set/alter fitting parameters page (Section 3.2.4)** (or that page can be reached by using the **Navigation tree (Section 3.1.4)**, or by right-clicking on the **Analysis window (Section 3.1.3)** to bring up a navigation context menu), and then clicking the check box to the right of the meniscus parameter to turn fitting back on, giving a page that looks like this:

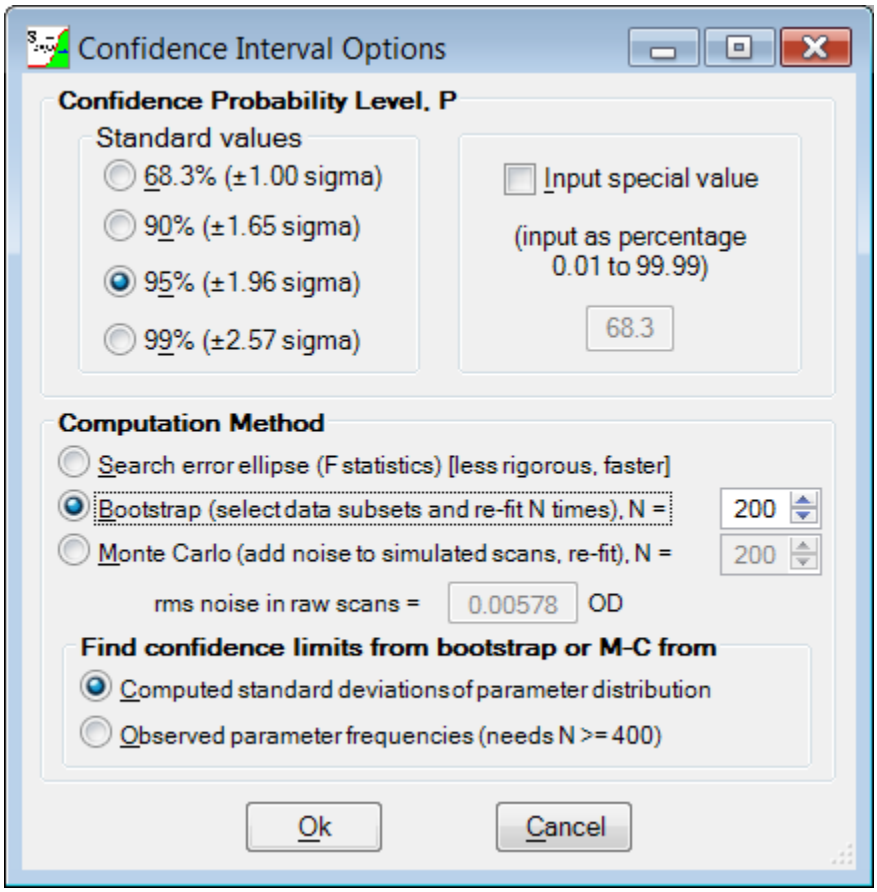


Then we move forward again and re-fit, producing this result (nearly the same as the previous one):


Analysis1
Fitted Parameters
Species
Parameter
Best Value
Estimated Standard Error
Starting V
1 M [kDa] 146.26 ± 0.18 146.26
2 fraction [%] 0.06 ± 0.02 0.06
2 s [S] 3.48 ± 0.06 3.48
3 fraction [%] 2.86 ± 0.06 2.86
3 s [S] 9.461 ± 0.013 9.461
4 fraction [%] 0.13 ± 0.10 0.12
4 s [S] 12.5 ± 0.5 12.4
5 fraction [%] 0.13 ± 0.06 0.14
5 s [S] 15.2 ± 0.3 15.2
- meniscus [cm] 5.90778 ± 0.00004 5.9077
View parameter cross-correlations
Compute confidence intervals
Create Reports
Include Graph
Print report
Send report to clipboard
Graph Monte Carlo results
Fixed, Constrained, & Implicit Parameters
Parameter or Constraint
Valu
cell base (cm) 7.200
slope factor (%/cm) 0.000
M constraint 1 M(2) = M(1) x [s(2)/s(1)]^
M constraint 2 M(3) = M(1) x [s(3)/s(1)]^
M constraint 3 M(4) = M(1) x [s(4)/s(1)]^
Fit statistics
Sum of squared residuals 1.743711E0
Degrees of freedom 51,762
RMS residual 5.774238E-3
Runs test Z 10.3
Log
Back
Next

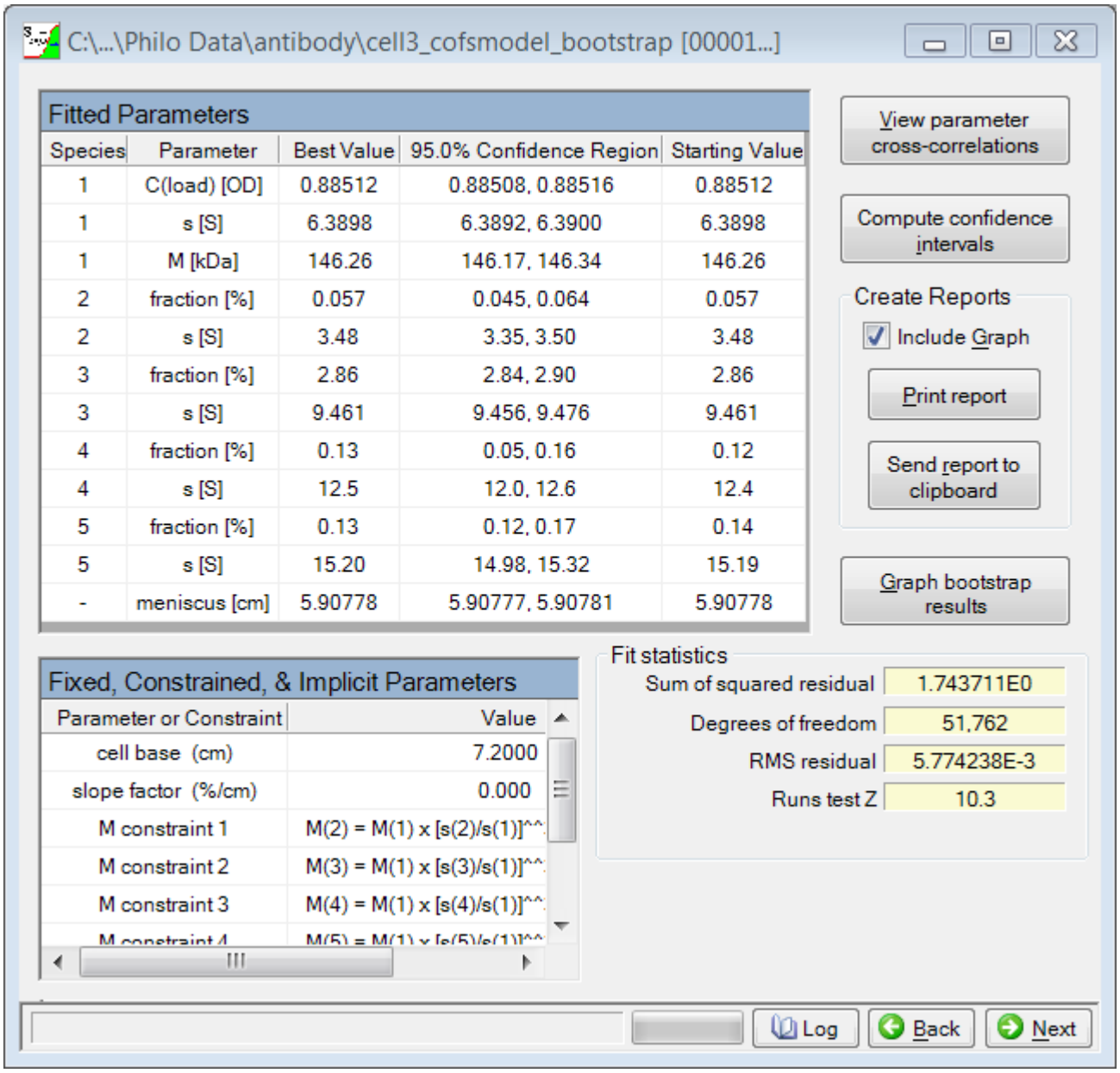
(10) At this point we now have a final fit, so this is a good time to save it to disk. We also have rough estimates for the precision of these parameters, and those are revealing. Note for example that the fraction of species 4 (probably a trimer) is 0.13% ± 0.10%, suggesting that this species may be only marginally non-zero. Similarly the estimated precision for its sedimentation coefficient is ± 0.5 S, the lowest precision among these 5 species.

However since one of the main advantages of transferring the *c(s)* analysis into this discrete species model is that we can derive true confidence intervals for all the parameters, you will probably want to do that. Pushing the **Compute confidence intervals button (Section 3.2.6.4)** brings up the dialog for selecting the desired confidence level and choosing the statistical approach for this calculation. In this case we will calculate 95% confidence intervals using the **Bootstrap method (Section 6.6)** (which has the advantage that it does not assume the noise in the raw data is random).



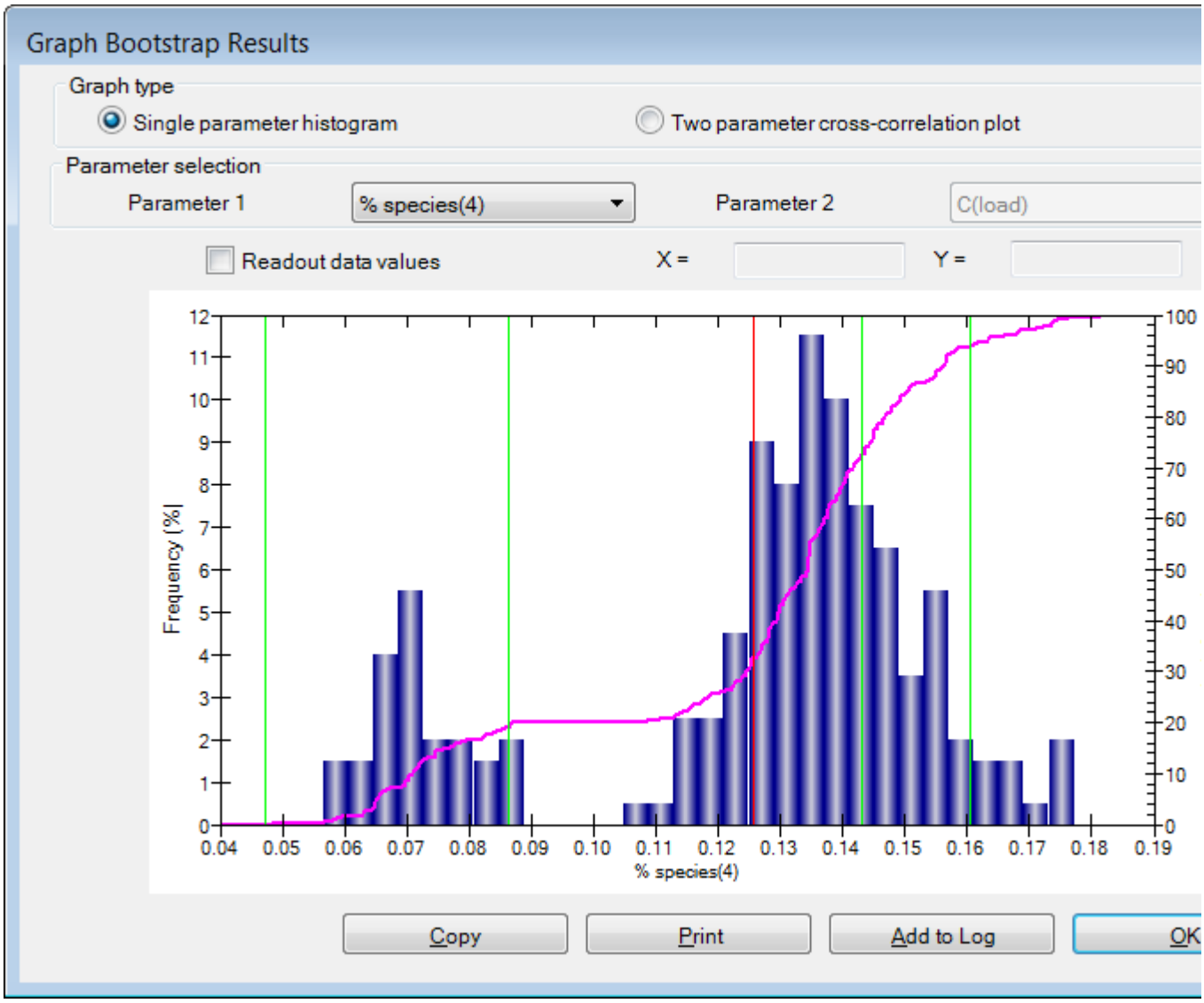
When there are many species or many interference scans are being fitted these calculations can take a considerable time (~50 min in this case). While it is desirable to use 200 or more rounds of bootstrap or Monte Carlo to define the limits, adequate precision can usually be obtained with 100 rounds (or even less). Computing the confidence limits using *F* statistics (the **Search error ellipse** option) is significantly faster, but that approach makes a number of assumptions about the statistical properties of the raw data.

The results of the 200 rounds of bootstrap are shown below:


C:\...\Philo Data\antibody\cell3_cofsmodel_bootstrap [00001...]
Fitted Parameters
Species | Parameter | Best Value | 95.0% Confidence Region | Starting Value
1 | C(load) [OD] | 0.88512 | 0.88508, 0.88516 | 0.88512
1 | s [S] | 6.3898 | 6.3892, 6.3900 | 6.3898
1 | M [kDa] | 146.26 | 146.17, 146.34 | 146.26
2 | fraction [%] | 0.057 | 0.045, 0.064 | 0.057
2 | s [S] | 3.48 | 3.35, 3.50 | 3.48
3 | fraction [%] | 2.86 | 2.84, 2.90 | 2.86
3 | s [S] | 9.461 | 9.456, 9.476 | 9.461
4 | fraction [%] | 0.13 | 0.05, 0.16 | 0.12
4 | s [S] | 12.5 | 12.0, 12.6 | 12.4
5 | fraction [%] | 0.13 | 0.12, 0.17 | 0.14
5 | s [S] | 15.20 | 14.98, 15.32 | 15.19
- | meniscus [cm] | 5.90778 | 5.90777, 5.90781 | 5.90778
View parameter cross-correlations
Compute confidence intervals
Create Reports
Include Graph
Print report
Send report to clipboard
Graph bootstrap results
Fixed, Constrained, & Implicit Parameters
Parameter or Constraint | Value
cell base (cm) | 7.2000
slope factor (%/cm) | 0.000
M constraint 1 | M(2) = M(1) x [s(2)/s(1)]^^
M constraint 2 | M(3) = M(1) x [s(3)/s(1)]^^
M constraint 3 | M(4) = M(1) x [s(4)/s(1)]^^
Fit statistics
Sum of squared residual | 1.743711E0
Degrees of freedom | 51,762
RMS residual | 5.774238E-3
Runs test Z | 10.3
Log
Back
Next

Comparison with the standard error estimates (1 σ, 68.3% confidence) above shows that in many cases the true precision is better than that estimate. However note that fairly often the confidence region is quite asymmetric around the best-fit value. For example, the best-fit sedimentation coefficient for species 2 is 3.48 S, but on the low side the confidence region is 0.13 S wide and on the high side it is only 0.02 S wide.

It can also be quite informative to graph the distribution of parameter values from the bootstrap or Monte Carlo simulations (the **Graph bootstrap results button (Section 3.2.6.8)**). This graph of the values for the fraction of species 4 shows the distribution actually shows two peaks, one centered near 0.07% and another near 0.14%. A number of other parameters also show double-peaked distributions.



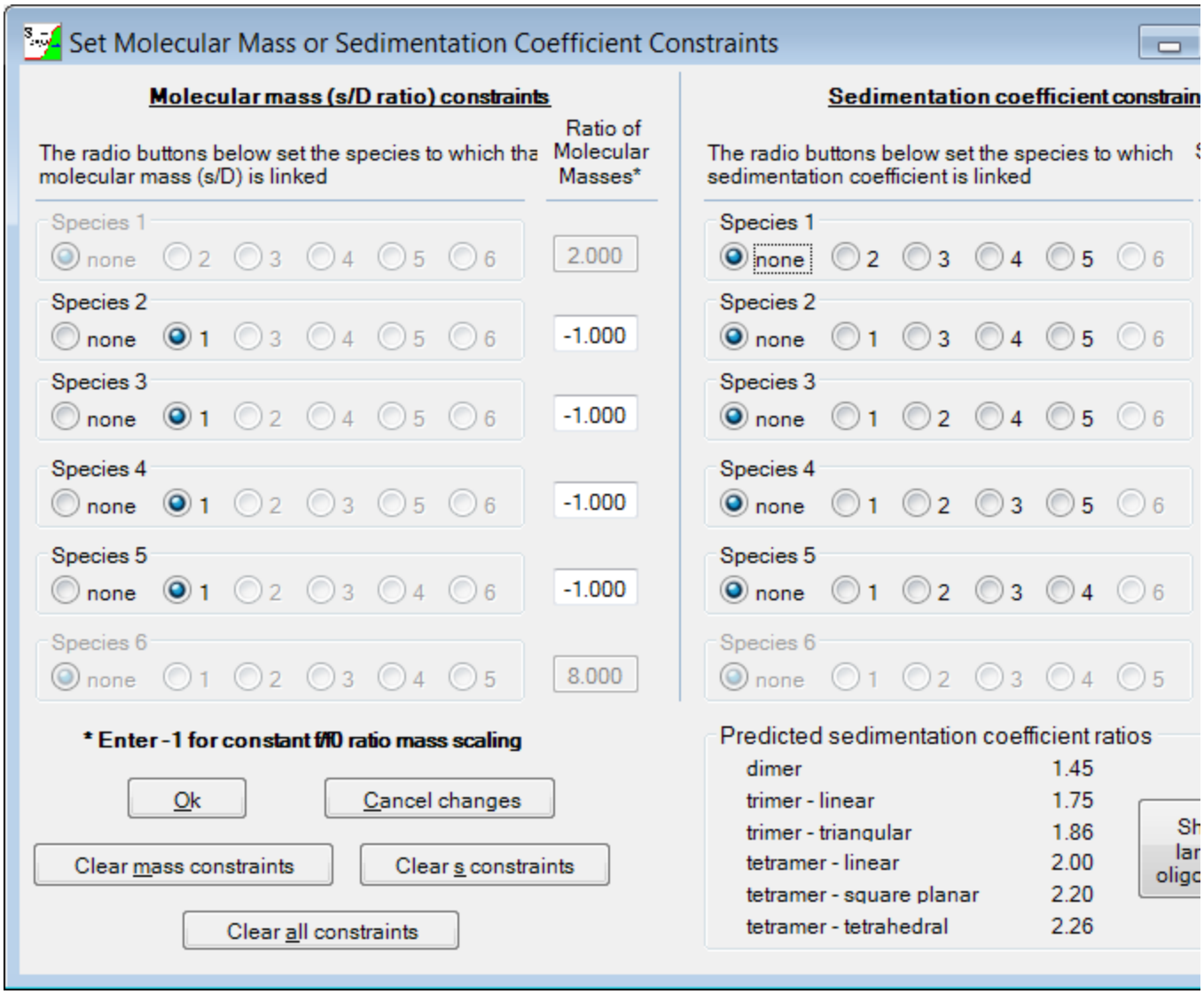
Thus the parameter probability distribution in this case clearly cannot be adequately described by a single symmetric Gaussian curve, as is often assumed!

Altering or removing the parameter constraints

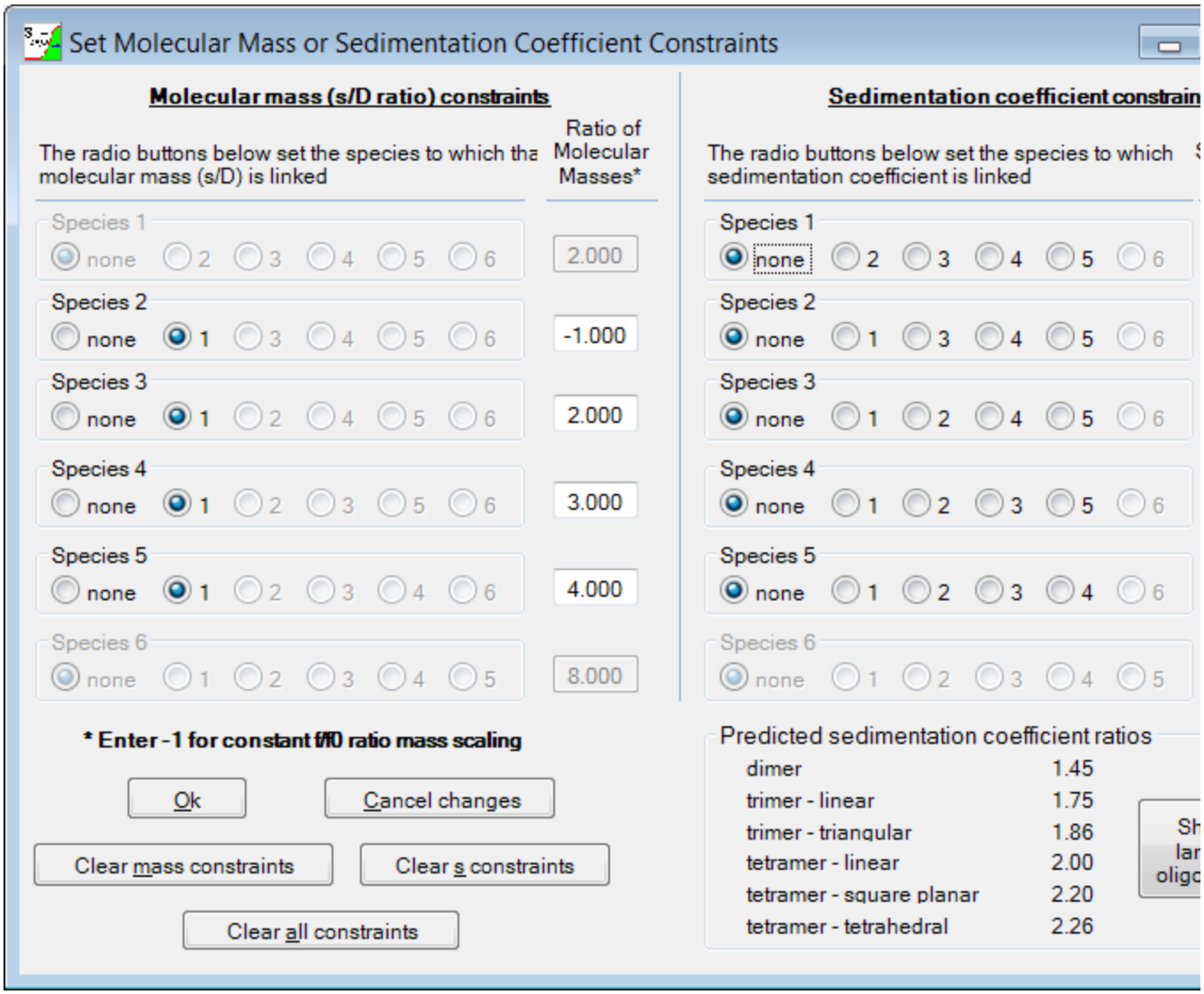
Another major advantage of transferring a *c(s)* fit into a discrete model is that one can explore alternatives to the assumption that all species have the same f/f_0 ratio, including testing whether it is possible to remove constraints and directly fit for the mass of some species (ones that represent a reasonably large fraction of the total). For many samples it may be reasonable to assume that certain species represent different oligomers of a common subunit, and thus their masses are related by integer ratios.

For this antibody example it is reasonable to assume that the species sedimenting faster than the main species (antibody monomer) represent oligomers, and the sedimentation coefficients of those minor peaks in the *c(s)* distribution seem consistent with assignments as dimer, trimer, and tetramer. In general oligomers will not have the same f/f_0 ratio as the monomer, and thus the *c(s)* model may actually be producing biased, inaccurate results due to the incorrect assumption of constant f/f_0 ratio.

Let's try changing the mass constraints to ones appropriate for dimer, trimer, and tetramer. To do this, return to the **Select fitting model page (Section 3.2.3)** and push the **Set constraint values button (Section 3.2.3.7)**. This brings up the constraints dialog, which looks like this for our converted *c(s)* fit:

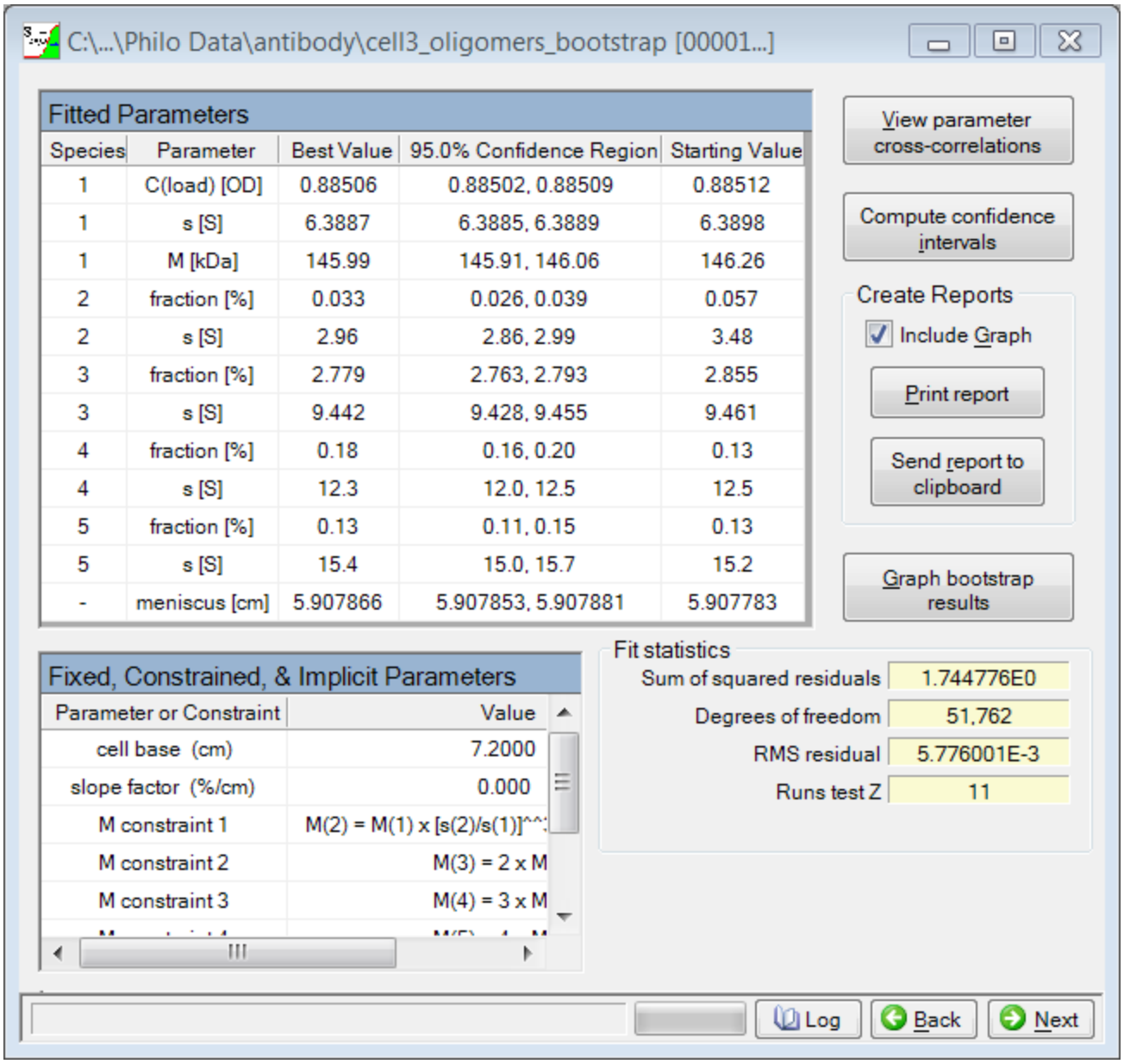

Set Molecular Mass or Sedimentation Coefficient Constraints
Molecular mass (s/D ratio) constraints
The radio buttons below set the species to which tha molecular mass (s/D) is linked
Ratio of Molecular Masses*
Species 1 none 2 3 4 5 6 2.000
Species 2 none 1 3 4 5 6 -1.000
Species 3 none 1 2 4 5 6 -1.000
Species 4 none 1 2 3 5 6 -1.000
Species 5 none 1 2 3 4 6 -1.000
Species 6 none 1 2 3 4 5 8.000
* Enter -1 for constant f/f0 ratio mass scaling
Ok
Cancel changes
Clear mass constraints
Clear s constraints
Clear all constraints
Sedimentation coefficient constrain
The radio buttons below set the species to which sedimentation coefficient is linked
Species 1 none 2 3 4 5 6
Species 2 none 1 3 4 5 6
Species 3 none 1 2 4 5 6
Species 4 none 1 2 3 5 6
Species 5 none 1 2 3 4 6
Species 6 none 1 2 3 4 5
Predicted sedimentation coefficient ratios
dimer 1.45
trimer - linear 1.75
trimer - triangular 1.86
tetramer - linear 2.00
tetramer - square planar 2.20
tetramer - tetrahedral 2.26

What we want to do is constrain species 3, 4, and 5 to be dimer, trimer, and tetramer of species 1. Since the masses of those species are already linked to that of species 1, all we need to do is change the entries in the mass ratio boxes from -1 (the value that indicates constant f/f_0 ratio scaling) to values of 2, 3, and 4 for species 3, 4, and 5 respectively. When set up for the oligomer model the dialog looks like this:


Set Molecular Mass or Sedimentation Coefficient Constraints
Molecular mass (s/D ratio) constraints
The radio buttons below set the species to which tha molecular mass (s/D) is linked
Ratio of Molecular Masses*
Species 1 none 2 3 4 5 6 2.000
Species 2 none 1 3 4 5 6 -1.000
Species 3 none 1 2 4 5 6 2.000
Species 4 none 1 2 3 5 6 3.000
Species 5 none 1 2 3 4 6 4.000
Species 6 none 1 2 3 4 5 8.000
* Enter -1 for constant f/f0 ratio mass scaling
Ok
Cancel changes
Clear mass constraints
Clear s constraints
Clear all constraints
Sedimentation coefficient constrain
The radio buttons below set the species to which sedimentation coefficient is linked
Species 1 none 2 3 4 5 6
Species 2 none 1 3 4 5 6
Species 3 none 1 2 4 5 6
Species 4 none 1 2 3 5 6
Species 5 none 1 2 3 4 6
Species 6 none 1 2 3 4 5
Predicted sedimentation coefficient ratios
dimer 1.45
trimer - linear 1.75
trimer - triangular 1.86
tetramer - linear 2.00
tetramer - square planar 2.20
tetramer - tetrahedral 2.26

Note that species 2 (a fragment) will be left with the constant f/f_0 ratio constraint---that may not be a correct assumption, but since it is present at a level below 0.1% it will be impossible to directly fit for the mass of this species (to leave its mass un-constrained).

After pushing the **Ok button** to confirm the changes to the mass constraints, the data can be re-fitted using the constant f/f_0 ratio model fit results as the initial guesses for this new model. The results of that fit and evaluation of the 95% confidence intervals using 200 rounds of bootstrap are shown below:


C:\...\Philo Data\antibody\cell3_oligomers_bootstrap [00001...]
Fitted Parameters
Species | Parameter | Best Value | 95.0% Confidence Region | Starting Value
1 | C(load) [OD] | 0.88506 | 0.88502, 0.88509 | 0.88512
1 | s [S] | 6.3887 | 6.3885, 6.3889 | 6.3898
1 | M [kDa] | 145.99 | 145.91, 146.06 | 146.26
2 | fraction [%] | 0.033 | 0.026, 0.039 | 0.057
2 | s [S] | 2.96 | 2.86, 2.99 | 3.48
3 | fraction [%] | 2.779 | 2.763, 2.793 | 2.855
3 | s [S] | 9.442 | 9.428, 9.455 | 9.461
4 | fraction [%] | 0.18 | 0.16, 0.20 | 0.13
4 | s [S] | 12.3 | 12.0, 12.5 | 12.5
5 | fraction [%] | 0.13 | 0.11, 0.15 | 0.13
5 | s [S] | 15.4 | 15.0, 15.7 | 15.2
- | meniscus [cm] | 5.907866 | 5.907853, 5.907881 | 5.907783
View parameter cross-correlations
Compute confidence intervals
Create Reports
Include Graph
Print report
Send report to clipboard
Graph bootstrap results
Fixed, Constrained, & Implicit Parameters
Parameter or Constraint | Value
cell base (cm) | 7.2000
slope factor (%/cm) | 0.000
M constraint 1 | M(2) = M(1) x [s(2)/s(1)]^^
M constraint 2 | M(3) = 2 x M
M constraint 3 | M(4) = 3 x M
Fit statistics
Sum of squared residuals | 1.744776E0
Degrees of freedom | 51,762
RMS residual | 5.776001E-3
Runs test Z | 11
Log
Back
Next

Although for the most part these results are not very different from those shown above for the constant f/f_0 ratio model, they reveal several interesting points:

- most of the 95% confidence intervals do not overlap with those from the constant f/f_0 ratio model
- the species whose properties show the largest changes is species 2 (the fragment), the one for which the mass constraint was not changed
- the confidence intervals for the oligomer model fit are much more symmetric, and none of the parameter value distributions show double peaks

Are these confidence intervals realistic?

The results from the oligomer model fit imply that the fractions of individual minor peaks can be determined within ±0.02%, and that the monomer sedimentation coefficient has a precision of ±0.003%. Given that we have fitted more than 50,000 data points, each with a signal/noise ratio of >150, it makes statistical sense that such precision is indeed possible, and yes these confidence limits are valid if the dominant noise sources in

the experiment are not systematic errors.

In practice it is extremely unlikely that multiple repeats of this experiment would give results falling within these calculated confidence limits, for a number of reasons. Temperature accuracy is an important source of variability in sedimentation coefficients. Convection is another major source of variability, and the alignment of the cells and quality of the centerpieces and cell housings is critically important for preventing convection (**reference 13 (Section 7.13)**).

Can the mass constraints be removed?

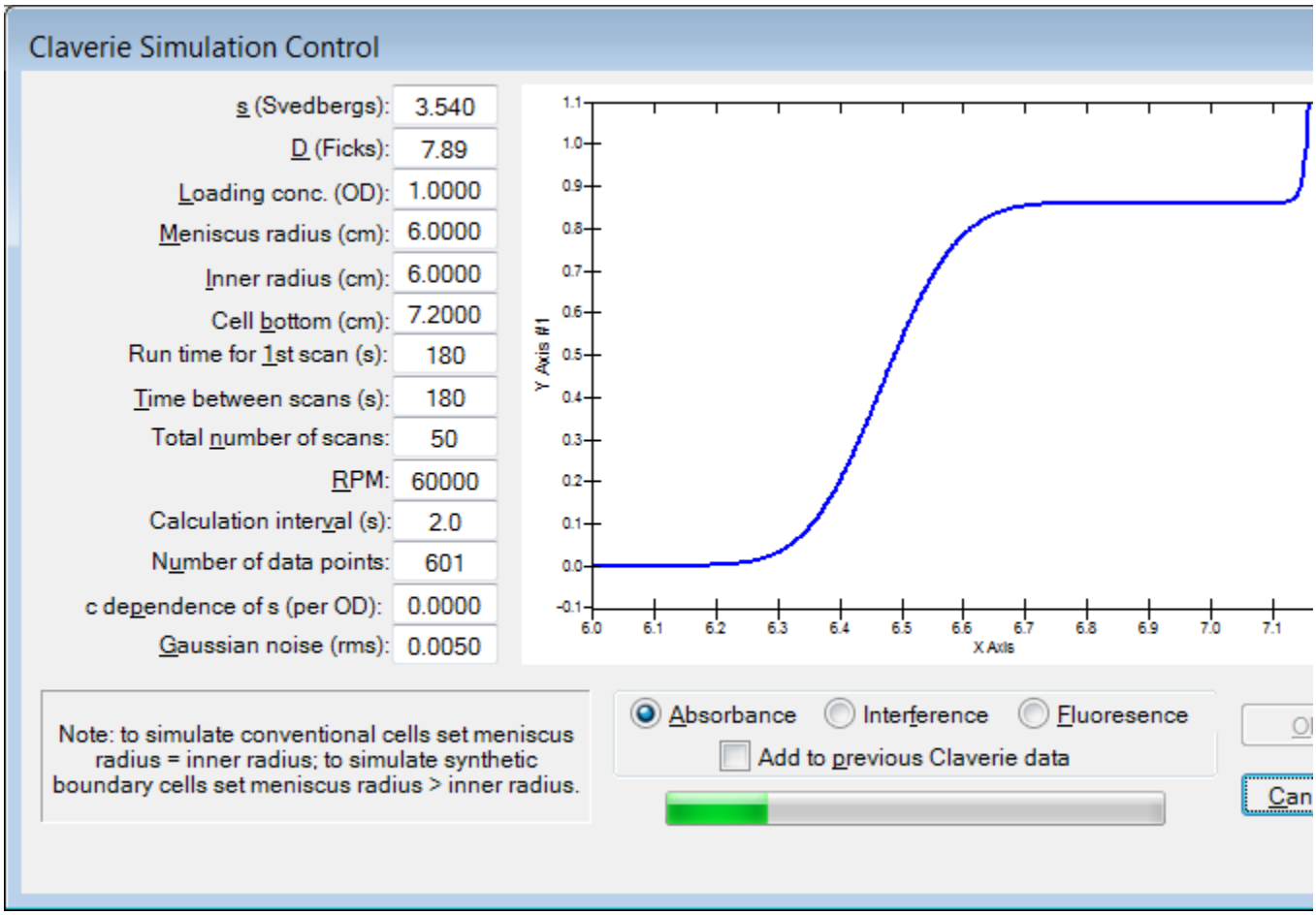
In some cases it will be possible to remove the mass constraint and thus directly determine the mass (and its confidence limits) at least for those species which represent a reasonably large fraction of the total. This is not true in the example we have been using thus far---removing the mass constraint even for species 3 (the most abundant minor species, at 2.8% of the total) and attempting a fit results in the mass of species 3 dropping toward the monomer mass, and simultaneously the fraction of species 4 drops to zero, which in turn prevents the fit from ever converging.

However for an example of a sample with low-level minor components where constraints are used initially and then can be removed see **Tutorial: Multi-species fits (Section 2.2)**.

2.4 Tutorial: Simulating experiments with the Claverie simulator

This tutorial first covers how to simulate a sample containing a single species, and then how to do multi-species simulations.

- **Step-by-step: single species simulation**


Claverie Simulation Control
s (Svedbergs): 3.540
D (Ficks): 7.89
Loading conc. (OD): 1.0000
Meniscus radius (cm): 6.0000
Inner radius (cm): 6.0000
Cell bottom (cm): 7.2000
Run time for 1st scan (s): 180
Time between scans (s): 180
Total number of scans: 50
RPM: 60000
Calculation interval (s): 2.0
Number of data points: 601
c dependence of s (per OD): 0.0000
Gaussian noise (rms): 0.0050
Y Axis #1
X Axis
Note: to simulate conventional cells set meniscus radius = inner radius; to simulate synthetic boundary cells set meniscus radius > inner radius.
Absorbance
Interference
Fluoresence
Add to previous Claverie data

1. Decide if you are simulating absorbance, interference, or fluorescence data and check the appropriate radio button (at the lower right on the form).
    - This choice affects the units of concentration, the default number of radial points, the allowed range for the loading concentration and added noise, and the file names for the generated data sets. For absorbance mode the default 601 data points corresponds to approximately the number of data points actually collected when using the default 0.003 cm point spacing for velocity scans. For interference and fluorescence modes the default 1641 and 601 data points, respectively, correspond approximately to the radial spacing of of those optical systems.

2. Decide if you are simulating conventional cells or a synthetic boundary cell. For a conventional cell the inner radius and the meniscus position are set to be equal; for a synthetic boundary cell the meniscus is at a radius greater than that of the cell inner radius.
    - Making the meniscus and inner radius equal for conventional cells essentially prevents diffusion of solutes to the region above the meniscus. In a synthetic boundary cell solutes can diffuse to radii inside the initial boundary position.

3. Enter the desired concentration, sedimentation coefficient, and diffusion coefficient for the species being simulated.
    - If you are trying to simulate a particular molecular mass, use the **Mass/diffusion calculator (C:\Users\Public\Documents\Visual Studio Projects\Svedberg Help Source\{8EABF9F3-CF47-45CE-80D1-1C75FF48FC72})** form to calculate the corresponding diffusion coefficient.

4. Enter the run time (elapsed time) for the first data set, the time interval between data sets, the total number of data sets (1000 maximum), and the rotor speed. (See **Tip** below.) For *dc*/*dt* analysis you will want the interval between sets to be fairly small (perhaps 60 s to simulate interference data or 240 s to simulate absorbance data in a 4-hole rotor).
5. Enter the calculation interval in seconds. This is the time increment between successive calculations of the concentration distribution.
    - Smaller intervals will give a more accurate, but slower, simulation. Faster moving boundaries will require smaller intervals; for low speeds or small sedimentation coefficients larger intervals are fine. A 1 sec interval should give good accuracy in most cases, but for sharp boundaries to get accurate boundary widths (diffusion coefficients) you may need to go to ~0.1 s. The minimum interval is 0.01 sec, maximum 50 sec.
    - **For highly accurate simulations of conditions giving sharp boundaries you will need a small interval and the simulation will likely require up to a minute to complete.** Absorbance simulations require smaller intervals because the radial spacing is larger (but they calculate faster at a given interval for the same reason).

6. Enter the number of data points to cover the radial range between the inner radius and cell bottom. Again, more points gives more accuracy but a longer calculation.
    - Sharper boundaries will require more radial points for accuracy. The data files that are generated will have data at these radial positions.

7. If desired, specify a nonzero value for the concentration dependence coefficient of the sedimentation coefficient.
    - This algorithm assumes a linear dependence of *s* with concentration, given by the formula $s = s_0(1 - kc)$, where *k* is the coefficient entered on the form. For globular proteins an empirical value for *k* is often taken to be 0.009 ml/mg (corresponding to *c* in mg/ml). In this program the units for *k* are per AU or per fringe. The *D* value has no concentration dependence.

8. If desired, specify a nonzero value for the random Gaussian noise to be added to the simulated curves.

9. Press the 'OK' button to start the simulation. The progress bar will show approximately how far the calculations have progressed, and they can be aborted using the 'Cancel' button.
10. When the simulation is done, the data are automatically loaded into the internal data arrays, just as though they had been read in as scan files from a disk. You will be asked whether you wish to also save these results to disk as scan files so they could be reloaded at a later date.
    - If you save the scans you may specify which folder to put them into using the File Save dialog box that appears. This dialog box also allows you to create a new folder in which to store the simulation. The scans are normally written in sequence starting with 00001. You will also be asked to provide a comment line for the scans. Also please note that in order to provide an obvious meniscus position for simulations of conventional cells, when the data are written to disk two extra data points to the left of the meniscus are added, and the data point at the meniscus position is set to a value equal to the loading concentration. This produces a 'spike' at the meniscus, similar to that seen in real absorbance data.
11. Finally, you will automatically be taken to the **Set meniscus & data region page (C:\Users\Public\Documents\Visual Studio Projects\Svedberg Help Source\{14FCF551-188C-4C48-9E48-0B1B6957B1E6})** to view a graph of the simulation you just created.
    - Note that in simulated data the concentration at the cell base often builds up to very high values. Thus the graph will be automatically zoomed to give a range from 0 to the loading concentration (which means the region where solutes accumulate will go off scale).

Finding the right scan interval and number of scans to get the proper range of boundary movement can take some trial and error. Do some quick, low accuracy, simulations with a long calculation interval to get the time range correct, and then do a final accurate simulation with a smaller time interval. If you are going to be summing data for more than one species, remember you will want to pick a run time that is appropriate for all the species.

Step-by-step: multi-species simulation

Multi-species simulations are done by summing single-species simulations. This feature is governed by the 'Add to previous Claverie data' check box near the bottom of the form. This check box becomes active when Claverie data are currently in memory. It is only possible to sum simulations that generate compatible data sets (same radii, start time, *etc.*). That is, in summing simulations the only parameters changed between species are their *s* and *D* values and concentration.

The steps for a multi-species simulation are:

1. Perform a simulation for the first species.
    - You may save the results to disk if you want to keep that data, but it is not necessary to save to disk.
2. Immediately after the first simulation, invoke the Claverie simulation command again.
3. Enter the desired concentration, sedimentation coefficient, and diffusion coefficient for the next species.
4. Be sure to check the 'Add to previous Claverie data' check box (it is automatically cleared when the form is loaded).
5. Run the simulation for the next species.
6. Note that the graph being displayed shows only the current simulation being generated, not the sum.
7. If there are additional species, return to step **2**.
8. After the last species has been calculated, you may wish to save the summed data to disk.

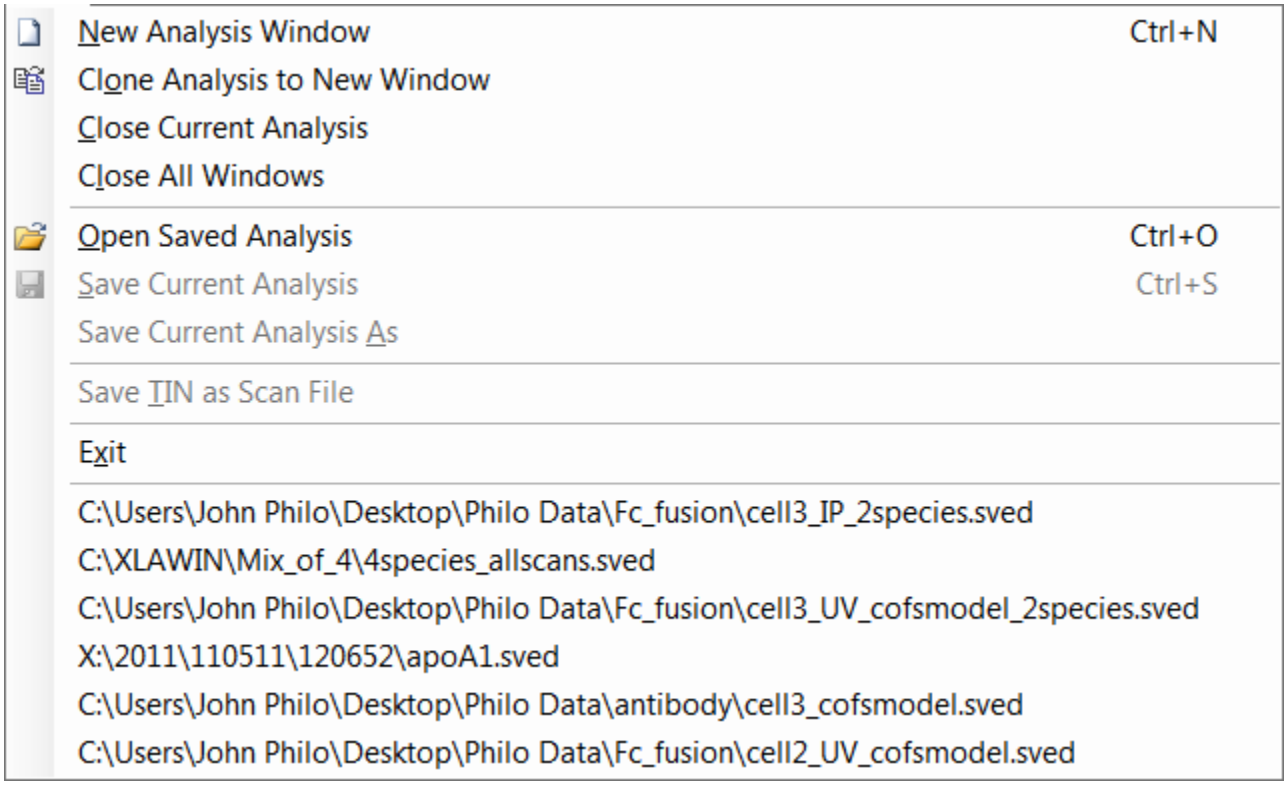
If you are going to add noise, you will probably not want to add it to each species, but rather add it only to the last species. If you added the same amount of noise to each species, the total r.m.s. noise will grow as the square root of the number of species.

3 Program elements, controls, and features

3.1 User interface

3.1.1 Main menus

3.1.1.1 File menu



New Analysis Window

start a fresh analysis in a new **Analysis window (Section 3.1.3)**. This menu command is duplicated by a button on the **Toolbar (Section 3.1.2)**. Shortcut = CTRL+N

Clone Analysis to New Window

copy the current analysis into a new analysis window so it can be modified without affecting the current window. This menu command is duplicated by a button on the **Toolbar (Section 3.1.2)**.

Close Current Analysis Window

close the active analysis window

Close All Windows

close all open analysis windows

Open Saved Analysis... (Section 3.1.1.1.1)

restores results and/or fits previously saved to disk. This menu command is duplicated by a button on the **Toolbar (Section 3.1.2)**. Shortcut = CTRL+O

Save Analysis (Section 3.1.1.1.2)

update analysis previously saved to disk. This menu command is duplicated by a button on the **Toolbar (Section 3.1.2)**. Shortcut = CTRL+S

Save Analysis As... (Section 3.1.1.1.3)

assign file name and save dc/dt and g(s*) curves and fitting results (if any) in a permanent record on disk

Save TIN as Scan File...

If the TIN has been computed for the current fit this menu item will be active. Saving the TIN as a scan file (with the headers and structure of a scan file, rather than just as columns of text) allows it to be displayed by any program that recognizes scan files or to be subtracted as a baseline. Selecting this item will bring up a **Save file dialog** allowing you to designate the file name and location (folder).

Exit

exits program after reminder to save unsaved or altered analyses to disk (if there are any)

(Recent Files)

list of 6 most recently saved or restored analysis files; click on these entries to restore that analysis (in a manner equivalent to the **Open Saved Analysis... (Section 3.1.1.1.1)** command)

3.1.1.1.1 Open saved analysis...

This command will restore the data, results, and states of various program variables fitting results that were previously saved using the **Save Analysis As... (Section 3.1.1.1.3)** command. This menu command is duplicated by a button on the **Toolbar (Section 3.1.2)**.

The command brings up a standard Windows® file dialog that allows the user to specify the file name and folder of a previously saved ".sved" file (saved by version 7 or later) or ".fit" file (saved by versions 1-6).

After the file is loaded, if the saved results include a completed fit, the program will bring up the **Report fit results page (Section 3.2.6)** to show the results from that fit.

The path to this restored fit will be added to the top of the recent files list at the bottom of the **File menu (Section 3.1.1.1)**.

Tip 1: Once you have opened a saved analysis, you may want to use the **Clone analysis to new window** command to create a second copy that can be altered while preserving the original analysis in an open window for reference.

Tip 2: Since the .sved file extension will be associated with this program, you can also restore a fit by double-clicking on the file name from Windows Explorer, or in the Recent Documents list on the Start Menu.

Note that when a saved analysis is loaded, the program attempts to set the default folder for scan file loading to the folder from which the data were originally loaded. If that folder is not currently available (for example if the data were on a removable drive or a network drive to which you are not currently connected) there is often a delay of several seconds while Windows tests for what drives and folders are currently available.

3.1.1.1.2 Save analysis

This command saves an updated version of an analysis that was previous saved using the **Save analysis as... (Section 3.1.1.1.3)** command. This command is enabled when the previously saved version is altered.

3.1.1.1.3 Save analysis as...

This command will write out a file that contains all the data (raw scans) and any fit results that have been obtained, up to and including the last calculation or fit that was performed. This provides a permanent record of your work that can be quickly reloaded if further processing is needed or to print graphs and reports, and to serve as an archival record of your work.

The command brings up a standard Windows® file dialog that allows the user to specify the file name and folder. The user can specify a filename of any length; the program will automatically add the extension of ".sved" to that name.

The data file is written in a proprietary binary format. All scans which have been loaded into memory are saved, independent of whether that scan is presently included in the analysis. The scan data require only 8 bytes of storage per data point (much more compact than the original ASCII scan files).

The path to this saved results file will be added to the top of the recent files list at the bottom of the **File menu (Section 3.1.1.1)**.

3.1.1.1.4 Clone analysis to new window

This command will open a new **analysis window (Section 3.1.3)** containing an exact copy of the current analysis window. That copy includes all the raw data (in memory, not the original scan files on disk), calculated *dc/dt* and *g(s*)* data, settings, *etc*. This menu command is duplicated by a button on the **Toolbar (Section 3.1.2)**.

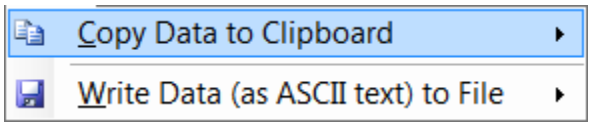
This allows you to alter the analysis while preserving an open copy of your original analysis. For example, you might want to change the number of species or how systematic noise is being handled. The name of the cloned window is derived by adding an underscore character and a sequence number to the name of the current analysis. Thus the first cloned version of a window called "cell2" would be "cell2_1".

An alternative way of altering an analysis while preserving a saved original version is to make the alterations and then save the new version under a new name using the **Save analysis as... (Section 3.1.1.1.3)** command. However with that approach you might accidentally overwrite the old version if you forget to alter the name.

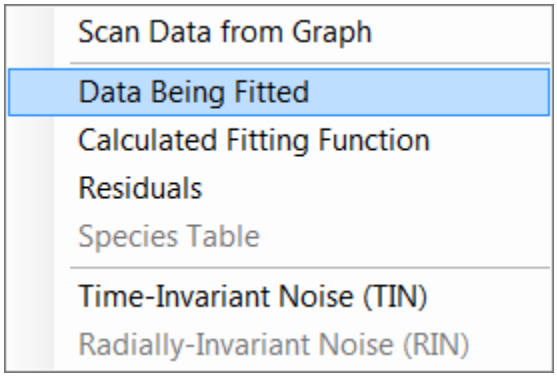
3.1.1.2 Edit menu

These commands export data from the current analysis as text tables. The data can either be copied to the Clipboard so these values can be pasted into some other program, or written to disk as text (ASCII) files.

These Copy or Write menu items under the Edit menu



lead to a submenu, as shown below, that determines which data to export:



When writing to disk, a standard File Save dialog allows you to specify a file name and the folder for the file. By default the file extension will be .dat, but you can also select .txt or any other extension.

When the data is exported the columns are separated by tab characters, and rows are ended by a carriage return and line feed.

The menu items below are only enabled after an initial fit (the starting point on the Perform least-squares fit page (Section 3.2.5)) or a fit simulation has been completed.

Scan data from graph

This menu selection will export the scan data (signal versus radius) from the graph shown on the **Set meniscus & data region page (Section 3.2.2)**. For interference scans this will give the data after removal of jitter and fringe jumps if that is what is currently shown on the graph. This menu item is only enabled when this graph is currently being displayed.

If the graph shows *N* scans then the data table will have 2*N* columns, with the first two columns holding the x and y values for the first scan, the 3rd and 4th columns holding the x and y values for the second scan, *etc*.

Note that copying the data into a large text string can take a minute or more when the number of scans is large. An hourglass is displayed while the text string is being created.

Because absorbance scans have different numbers of data points some of the columns in the data table will have fewer entries than others. Table positions with non-existent data are copied as empty strings.

*dc*Data being fitted

This menu selection will copy the actual data values being fitted (x,y pairs).

Calculated fit

This menu selection will copy the theoretical values from the fitted functions (x,y pairs).

Residuals

This menu selection will copy the residuals from the fit (x,y pairs).

Species table

This menu item is only enabled if 2 or more species were fitted. This menu selection will copy a table with 3 + N columns, where N is the number of species, with columns as follows:

1. an integer indicating which scan that row applies to

2. the radial position
3. the total fitted signal at that radial position
4. columns giving the signal contributions from each individual species (species1, species2, ...).

The data corresponding to different scans follow one after another (successive rows) so the total number of rows can be very large.

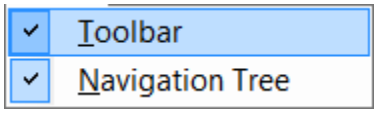
time-invariant noise (TIN)

This menu item is only enabled when fitting of the TIN is included in the fitting model (see **Select fitting model - Baseline fitting options (Section 3.2.3.5)**). It will copy the computed TIN (2 columns with pairs of radius and signal values).

Radially-invariant noise (RIN)

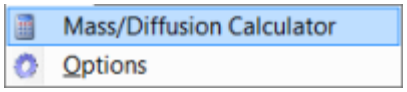
This menu item is only enabled when fitting of the RIN is included in the fitting model (see **Select fitting model - Baseline fitting options (Section 3.2.3.5)**). It will copy the computed RIN (1 column with vertical displacements relative to the first scan). The number of rows will be one less than the total number of scans being fitted.

3.1.1.3 View menu



The **View menu** allows you to toggle on or off the display of the **Toolbar (Section 3.1.2)** and the **Navigation Tree (Section 3.1.4)**. Turning off the **Toolbar (Section 3.1.2)** and/or the **Navigation Tree (Section 3.1.4)** will provide additional screen space for display of multiple analysis windows.

3.1.1.4 Tools menu



The **Tools menu** contains submenus to bring up either of two tools:

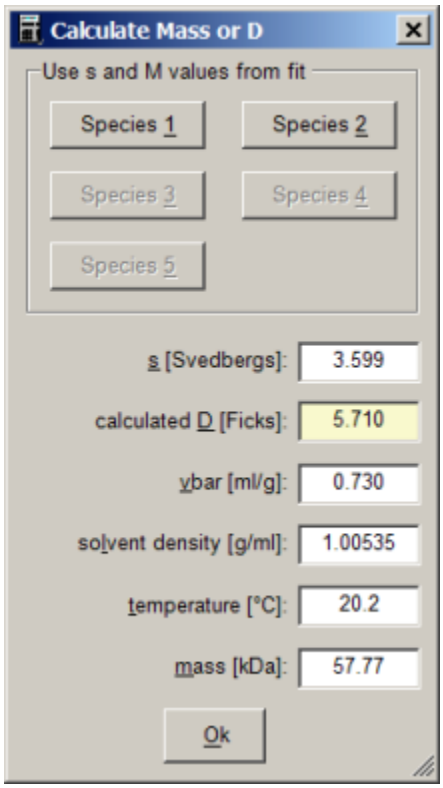
1. the **Mass/diffusion calculator (Section 3.1.1.4.1)** for calculating masses when the sedimentation and diffusion coefficients are known, or diffusion coefficients when the mass and sedimentation coefficient are known
2. the **F statistics calculator (Section 3.1.1.4.2)** for calculating statistical values related to confidence intervals or for comparing two different fits
3. the **Options dialog (Section 3.1.1.4.3)** for setting user preferences (defaults) for various program options

Each of these menu items is mirrored by an icon on the **Toolbar (Section 3.1.2)**.

3.1.1.4.1 Mass/diffusion calculator

This View Menu command or its equivalent button on the **Toolbar (Section 3.1.2)** will bring up a small calculator-like form (shown below) which

1. **calculates solute molecular masses** given values for sedimentation and diffusion coefficients, solute partial specific volume, solvent density, and temperature.
2. **calculates diffusion coefficients** if the mass and sedimentation coefficient are known.



This form is a handy way to determine the molecular mass corresponding to the *s* and *D* parameters from a fit where *D* is used as the fitting parameter, or to find the *D* value corresponding to a molecular mass when *M* has been used as a fitting parameter, without changing the **fitting model (Section 3.2.3)** and re-fitting.

How it works

This dialog box implements the Svedberg equation:

$$M = \frac{sRT}{D(1-\bar{v}\rho)}$$

where *R* is the gas constant, $\bar{v}$ is the partial specific volume, ρ is the solvent density, and *T* is temperature (in Kelvin).

The user may simply enter any desired values and the corresponding molecular mass will be calculated. When numbers are manually entered in the molecular mass box, then the corresponding *D* is calculated (and the label of this *D* value changes to 'calculated *D*'.

Species buttons

After a fit has been completed, a number of buttons at the top of the form will become active to correspond to

the number of species in that fit (2 in the case shown above). Pressing these buttons will fill in the parameters from that fit. The partial specific volume value associated with that species will also be used if that has been filled in on the **(Section 3.2.3)**select fitting model page.

Copy button

The 'Copy all values to Clipboard' button will copy the labels and values for these entries as a small table to the clipboard, with the labels and values separated by tab characters and a separate line for each type of entry.

About temperature

The program automatically fills in the sample temperature based on the temperature recorded in the header of the last scan used in calculating the *dc/dt* data, but since the temperature may drift during the run you may wish to adjust that value manually. The program internally converts the temperatures from Celsius to Kelvin scales.

Please note that changing the temperature on this form overrides the temperature recorded with the scans, and also alters the value used elsewhere in the program (for example, for conversions from raw *s* to *s(20,w)*).

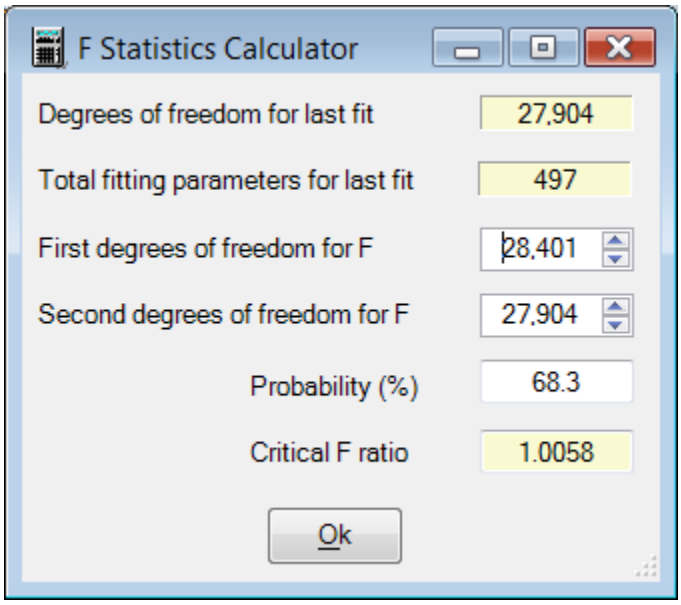
About s(20,w) units

If conversion to *s(20,w)* units is currently active, the sedimentation and diffusion coefficients will be *s(20,w)* and *D(20,w)* values, the vbar value will be value at 20 °C, the solvent density will automatically be set to that of water at 20 °C (and cannot be altered), and the temperature will be set to 20 °C (and cannot be altered). While conversion to *s(20,w)* units is active, changing the vbar on this form will not alter the vbar value used elsewhere in the program.

3.1.1.4.2 F statistics calculator

This item on the **Tools menu (Section 3.1.1.4)** brings up this small dialog box used for calculating the statistical *F* function, whose values can be used in evaluating parameter confidence intervals or for comparing whether on fit is significantly better than another. It is provided as a convenience---it is not needed for normal operations of this program such as evaluating confidence intervals.

F*(ν_1, ν_2, 1 - *p*)** is the *F* ratio (a ratio of two *chi squared* values) at which there is a probability ***p that two chi squared values are significantly different for ν_1 and ν_2 degrees of freedom. This *F* function which can be evaluated using this calculator tool by entering the values for ν_1 and ν_2 and ***p***.



By default if the current analysis window has a completed fit then the number of data points in that fit is entered as the first degrees of freedom and the *degrees of freedom* for that fit is entered as the second degrees of freedom, and the critical F ratio corresponding to a 68.3% confidence probability for those values is calculated and displayed, as shown above. Note that the number of fitting parameters is high for the example shown because the TIN was included in the fit.

Use in evaluating confidence limits

To evaluate the confidence limits with probability ***p*** for a single parameter from a fit of ***N*** total raw data points using a total of ***M*** fitting parameters and ν degrees of freedom one can manually increase or decrease the value of that parameter away from the best-fit value until the *chi squared* increases from the best fit value, χ^2_{min}, up to the limit that corresponds to that statistical probability (the so-called 'critical chi squared' value, χ^2_c). That is, the parameter values at which chi squared increases to χ^2_c define the limits of the confidence interval.

But how is χ^2_c determined? According to Equation 35 in **Johnson and Straume (1994) (Section 7.11)**, and also **Johnson & Faunt (1992) (Section 7.2)**, the correct formula for calculating the critical chi squared value that corresponds to a parameter confidence interval at probability ***p*** is

$$\chi^2_c = \chi^2_{min} \times \left(1 + \frac{M}{\nu} \times F(M, \nu, 1-p)\right)$$

This calculator will evaluate the last term in this equation (the *F* function) and values for the remaining terms are reported on this calculator dialog or the **Report fit results page (Section 3.2.6)**.

These equations are used internally by SVEDBERG in calculating confidence limits via *F* statistics (see **Compute confidence intervals button (Section 3.2.6.4)**), but rather than re-fitting as a parameter is varied instead the search of parameter space is made along axes of a hyper-ellipse that represent eigenvectors of the covariance matrix (following **reference 2 (Section 7.2)**).

Use for comparing different fits

F statistics can also be used for testing whether two different fits of the same data (for example fits with different numbers of species) are producing significantly different chi squared values. In this case the relevant ratio is the ratio of *reduced chi squared* values, and the formula used is (Equation 45 from **reference 2 (Section 7.2)**):

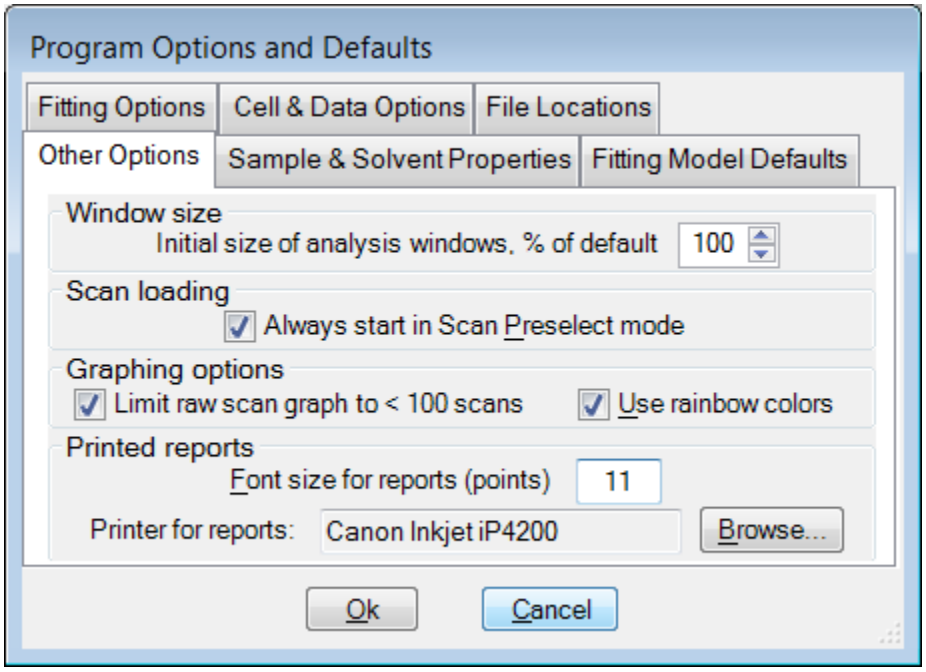
$$\frac{\chi_1^2/\nu_1}{\chi_2^2/\nu_2} = F(\nu_1, \nu_2, 1-p)$$

3.1.1.4.3 Options dialog

The **Options** item on the **Tools menu (Section 3.1.1.4)** will bring up the **Program Options dialog box** which uses multiple tabs to list various user preferences for program defaults. The options on each tab will be described below, with the default setting at program installation noted in square brackets [].

The **Cancel button** will close the dialog without altering any settings. The **Ok button** will save any new settings (which in general apply only to newly-opened analysis windows).

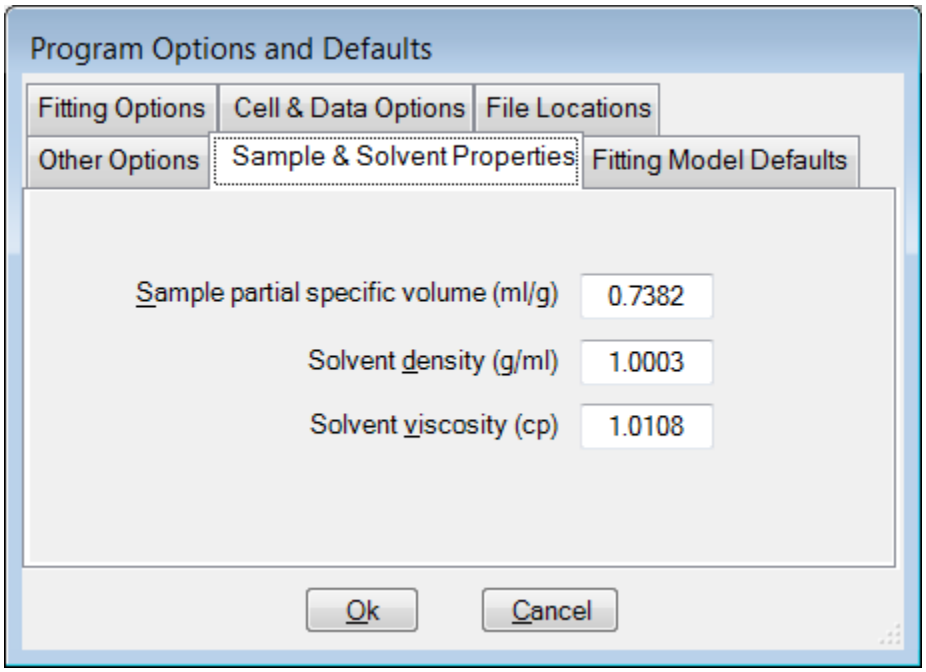
- **Other options tab**



- Initial size of analysis windows [default 100%] - this sets the size of newly-created analysis windows and other program dialog boxes as a percentage of the program default size. Smaller sizes will allow more analysis windows to be open without crowding, larger sizes will enhance readability (font size increases with window size). Allowed range 50%-150% of normal.
- Always start in **Scan Preselect mode (Section 3.2.1.1)** [default checked] - if checked then the default will be to pre-select the scan data type and cell number before opening the file dialog to select scans for loading.

- Font size for reports [12] - the font size, in points, to be used in printed reports of fitting results. Report text is always printed in the Arial font.
- Limit raw scan graph to < 100 scans [default checked] - if checked then the number of scans displayed on the **Set meniscus & data region page (Section 3.2.2)** will be limited to 100 or less by automatically setting the **Show every Nth scan** option on the **Graph display options tab (Section 3.2.2.1)**
- Use rainbow colors [default checked] - if checked then for graphs showing more than 7 traces the colors of the traces will be varied to form a rainbow pattern from violet to red rather than cycling through a set of 14 standard colors.
- Printer for reports [Windows default printer] - The text box shows the name of the current printer to be used for reports of fitting results. To alter this default use the **Browse...** button to open a Print dialog box and select from the drop-down list of available local or network printers.

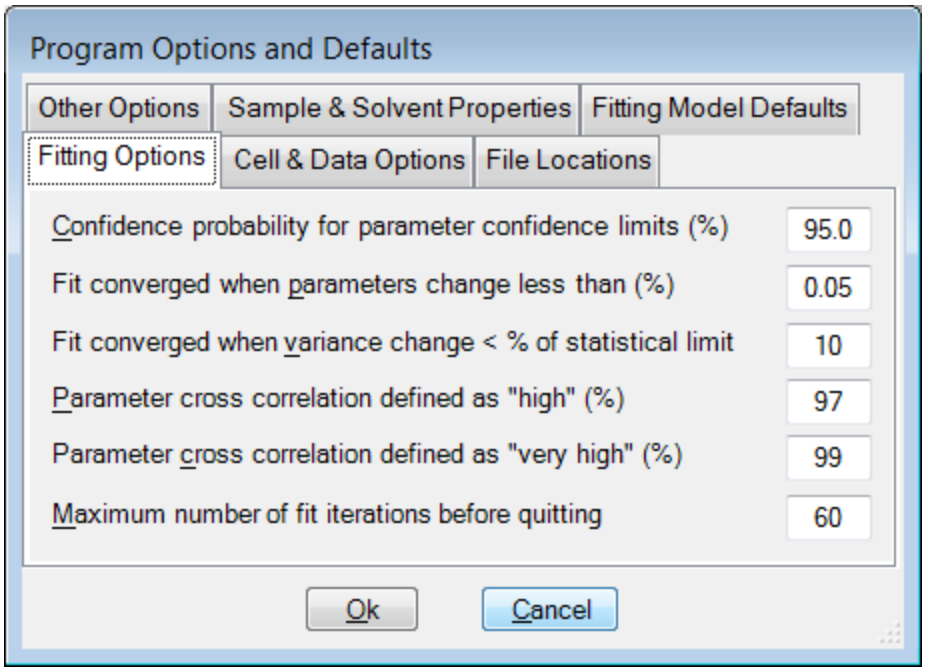
Sample & Solvent Properties tab



- Sample partial specific volume [0.73 ml/g]
- Solvent density [0.998234 g/ml] - the default is the density of water at 20 °C
- Solvent viscosity [1.002 cp] - the default is the viscosity of water at 20 °C. Note that these values are in centipoise, not poise as used in SEDNTERP and SEDFIT.

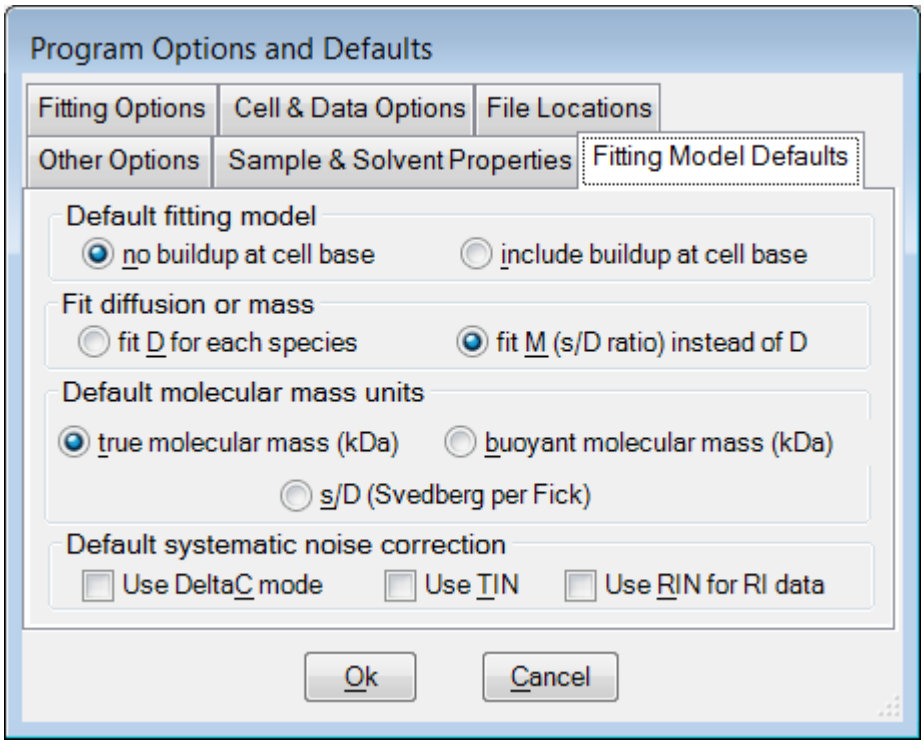
These default values are used for the first analysis window opened each time the program is started. When subsequent analysis windows are created, the last values that were entered in any previous window are used, which saves time when you are analyzing multiple samples in the same buffer or of the same protein.

Fitting Options tab



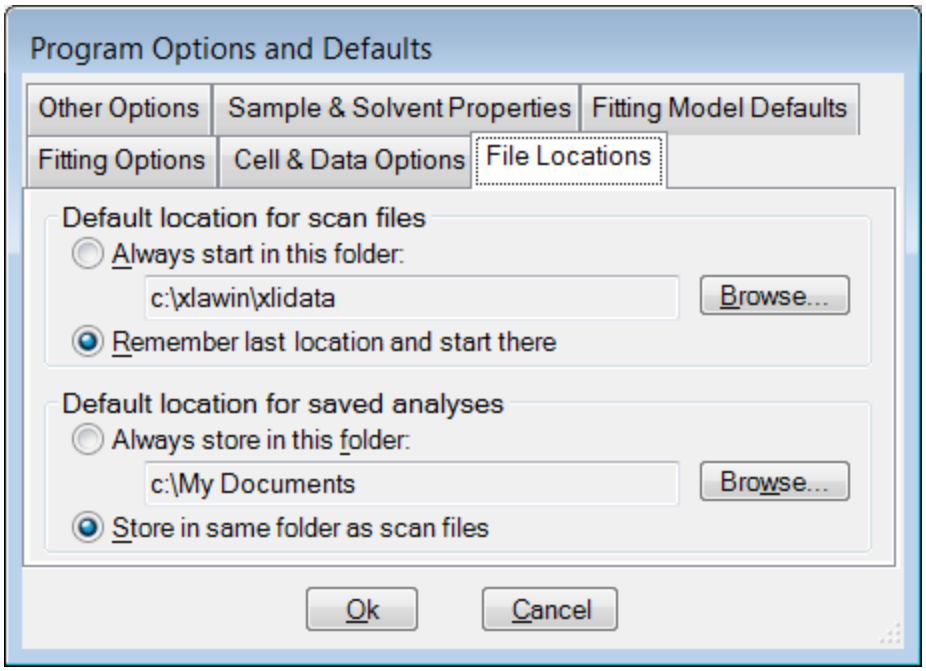
- Confidence probability for parameter confidence limits [68.3%] - This sets the default confidence level for parameter confidence limits. **Usually this is the only setting on this tab the user will want or need to alter.**
- Fit converged when parameters change less than [0.05%] - The fit iterations will halt when the last iteration changes all parameters less than this amount. See **Convergence control (Section 3.2.4.4)**.
- Fit converged when variance change < % of statistical limit [10%] - This provides a second criterion for fit convergence, halting the iterations when the improvement of the fit is no longer statistically significant. See **Convergence control (Section 3.2.4.4)**.
- Parameter cross-correlation defined as "high" [97%] - cross-correlations between this level and the "very high" setting will produce a mild warning on the **Report fit results page (Section 3.2.6)** and in printed reports. See **What does high parameter cross-correlation mean? (Section 5.1.6)**
- Parameter cross-correlation defined as "very high" [99%] - cross-correlations above this setting will produce a stronger warning on the **Report fit results page (Section 3.2.6)** and in printed reports. See **What does high parameter cross-correlation mean? (Section 5.1.6)**
- Maximum number of fit iterations before quitting [60] - This defines the maximum number of iterations before the fit halts with a warning message. Generally if a fit has not converged within 60 iterations it probably never will converge, and this limit is to prevent infinite or near-infinite loops.

Fitting model defaults tab



- No buildup / Include buildup at cell base [include buildup] - this default controls whether the default **fitting function (Section 3.2.3.3)** includes the buildup of macromolecules at the base of the cell (the 'Behlke function w/base') or treats the cell as semi-infinite and ignores the cell base (the 'Behlke boundary function).
- Fit diffusion or mass [fit M] - this default governs whether to fit the diffusion coefficient for each species or the species molecular mass (the *s/D* ratio). See **Select fitting model - Fit diffusion or mass (Section 3.2.3.6)**.
- Molecular weight units [true molecular mass] - this default sets the units to be used for molecular mass. See **Select fitting model - Molar mass units (Section 3.2.3.8)**.
- Systematic noise correction [default all off] - these check boxes allow you to, if desired, choose a default method for handling systematic baseline noise (either **delta-c mode (Section 3.2.3.4)**, a **fitted TIN (Section 3.2.3.5)**, or neither as the default) and also whether by default to handle *jitter* in Rayleigh interference data through a **fitted RIN (Section 3.2.3.5)**.

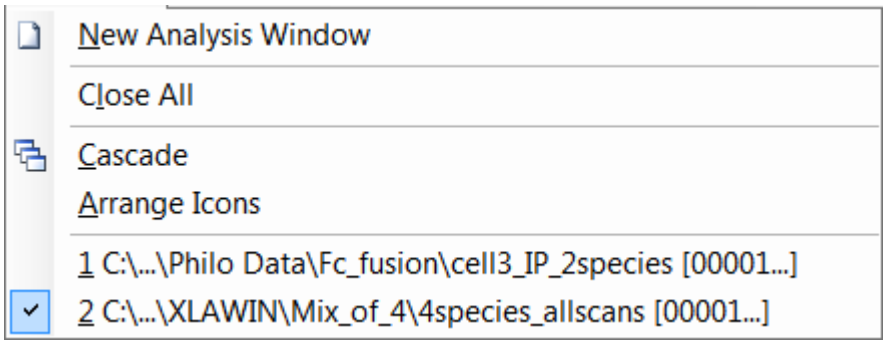
File locations tab



- Default location for scan files [remember last location] - These options govern which folder is initially displayed in the File Open dialog when new scan files are loaded.
  - The "Always start in this folder" option specifies that scan browsing always begins in a particular specified folder. For example, this might be a network drive, or a folder where all experiments for an individual XL-I user are stored. When this option is selected the Browse button is active and may be used to select the default folder.
  - The "Remember last location and start there" option has the program remember the last folder from which scans were loaded and start there.
- Default location for saved analyses [store with scans] These options govern which folder is initially displayed in the File Save dialog during the Save Analysis As... procedure.
  - The "Always store in this folder" option specifies that the File Save dialog always begins in a particular specified folder, such as the user's "My Documents" folder. When this option is selected the Browse button is active and may be used to select the default folder.
  - The "Store in same folder as scan files" option has the File Save dialog start in the folder from which the last scans were loaded. That approach has the advantage of keeping the saved analysis together with the raw data.

3.1.1.5 Windows menu

The **Windows menu** provides control over the display of the **analysis windows (Section 3.1.3)**.



New Analysis Window

This will create a new empty **analysis window (Section 3.1.3)**, and is equivalent to the File...New in the **File menu (Section 3.1.1.1)** or the New button on the **Toolbar (Section 3.1.2)**.

ose All

This provides a quick way to close all analysis windows (prompting you to save any containing unsaved results), freeing up screen space for new analyses.

Cascade

This is the usual Windows command to make all the open windows have a uniform size and arrange them is an overlapping cascade.

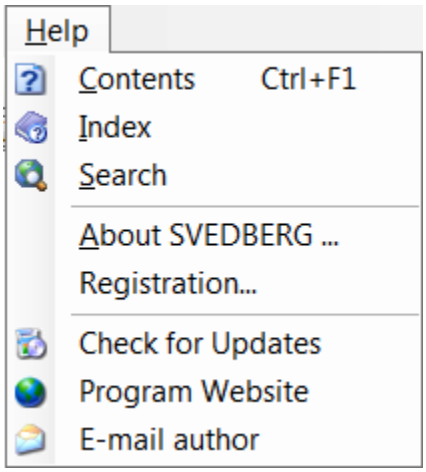
Arrange Icons

This is the usual Windows command that arranges the icons corresponding to any minimized windows along the bottom of the document area.

(Window List)

The bottom of this menu contains a list of all open windows, with a check mark next to the active window. You can activate a different window and bring it to the front by clicking its name, or by pressing the corresponding number on the keyboard.

3.1.1.6 Help menu



Contents

This brings up the Table of Contents tab for this Help file, and opens the **Program Overview and Organization (Section 1.1)** page.

Index

This brings up the Index tab for locating specific topics in the Help file.

Search

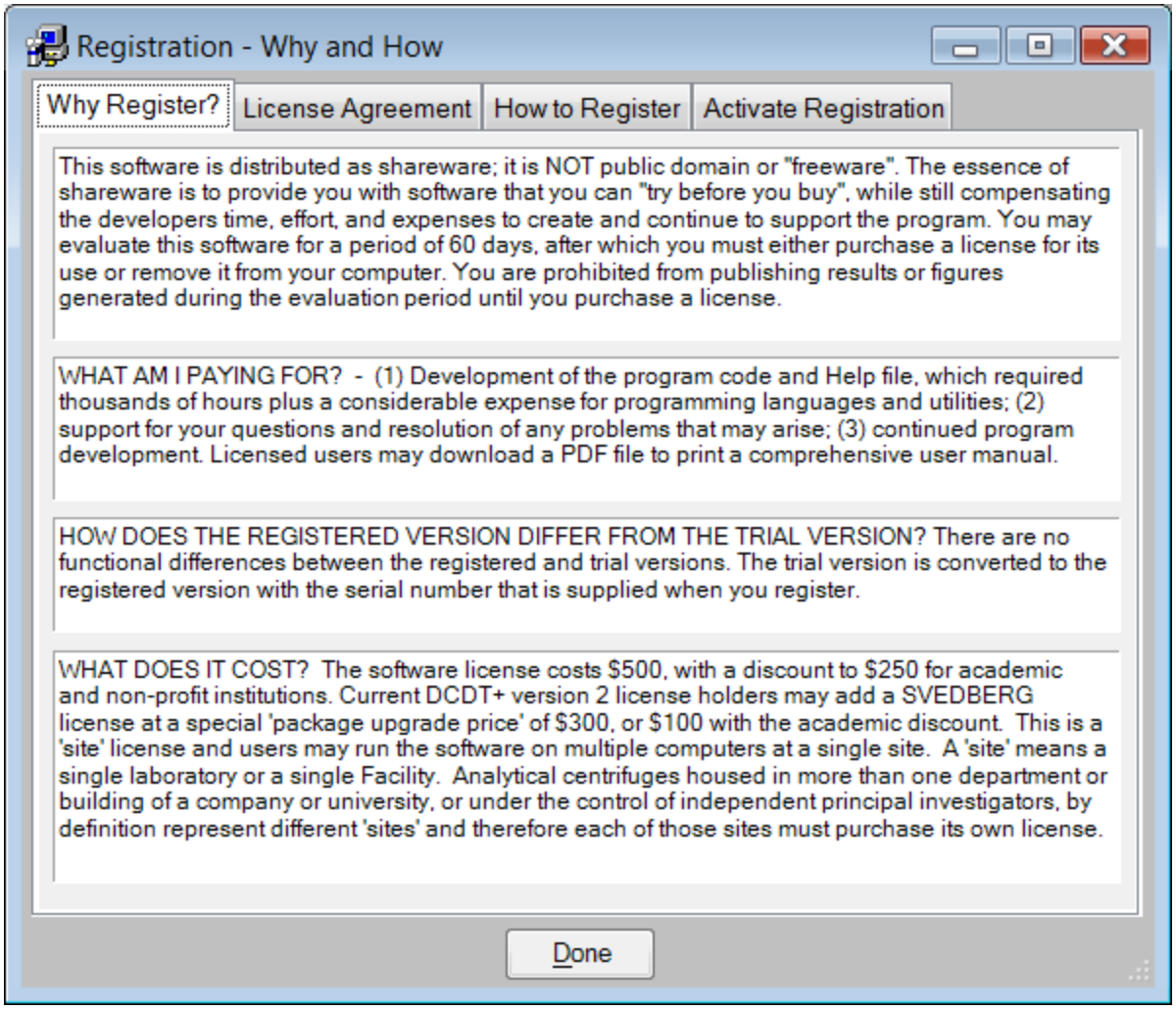
This brings up the Search tab for locating specific words or phrases that are not in the Index.

About SVEDBERG...

This brings up a dialog box showing the current version, registration information, and how to contact the author.

Registration

This brings up a tabbed dialog box with information about why and how to register the program, a page for entering and validating your program serial number, and buttons to display the software license agreement or print an order form.


Registration - Why and How
Why Register?
License Agreement
How to Register
Activate Registration
This software is distributed as shareware; it is NOT public domain or "freeware". The essence of shareware is to provide you with software that you can "try before you buy", while still compensating the developers time, effort, and expenses to create and continue to support the program. You may evaluate this software for a period of 60 days, after which you must either purchase a license for its use or remove it from your computer. You are prohibited from publishing results or figures generated during the evaluation period until you purchase a license.
WHAT AM I PAYING FOR? - (1) Development of the program code and Help file, which required thousands of hours plus a considerable expense for programming languages and utilities; (2) support for your questions and resolution of any problems that may arise; (3) continued program development. Licensed users may download a PDF file to print a comprehensive user manual.
HOW DOES THE REGISTERED VERSION DIFFER FROM THE TRIAL VERSION? There are no functional differences between the registered and trial versions. The trial version is converted to the registered version with the serial number that is supplied when you register.
WHAT DOES IT COST? The software license costs $500, with a discount to $250 for academic and non-profit institutions. Current DCDT+ version 2 license holders may add a SVEDBERG license at a special 'package upgrade price' of $300, or $100 with the academic discount. This is a 'site' license and users may run the software on multiple computers at a single site. A 'site' means a single laboratory or a single Facility. Analytical centrifuges housed in more than one department or building of a company or university, or under the control of independent principal investigators, by definition represent different 'sites' and therefore each of those sites must purchase its own license.
Done

Check for updates

This opens the Download page in your web browser.

Program web site

This opens the main SVEDBERG web page in your web browser.

E-mail author

This opens an e-mail for your questions, comments, or suggestions.

3.1.2 Toolbar

The **Toolbar** provides buttons as convenient alternatives for some of the menu commands.

The usual New, Open, and Save buttons are equivalents of the **New Analysis Window**, **Open saved analysis... (Section 3.1.1.1.1)**, and **Save analysis as... (Section 3.1.1.1.3)** items in the **File menu (Section 3.1.1.1)**. (Once an analysis has been saved and assigned a file name, the **Save button** becomes equivalent to the **Save analysis (Section 3.1.1.1.2)** command rather than **Save analysis as.... (Section 3.1.1.1.3)**)

The **Clone button** will copy the current analysis into a new window so it can be modified without affecting the current window.

The **Calculator button** is equivalent to the **Mass/diffusion calculator (Section 3.1.1.4.1)** entry in the **Tools menu (Section 3.1.1.4)** and pops up the Calculate Mass or D dialog box.

The **Options button** is equivalent to the **Options dialog (Section 3.1.1.4.3)** entry in the **Tools menu (Section 3.1.1.4)** and pops up the Program Options dialog box for setting user preferences.

The **Help button** is equivalent to the **Index** entry in the **Help menu (Section 3.1.1.6)** and will bring up the index for the program Help file.

> The display of the **Toolbar** is optional. It can be turned off to provide more screen area for **(Section 3.1.3)**Analysis windows using an entry under the **(Section 3.1.1.3)View menu (Section 3.1.1.3)**.

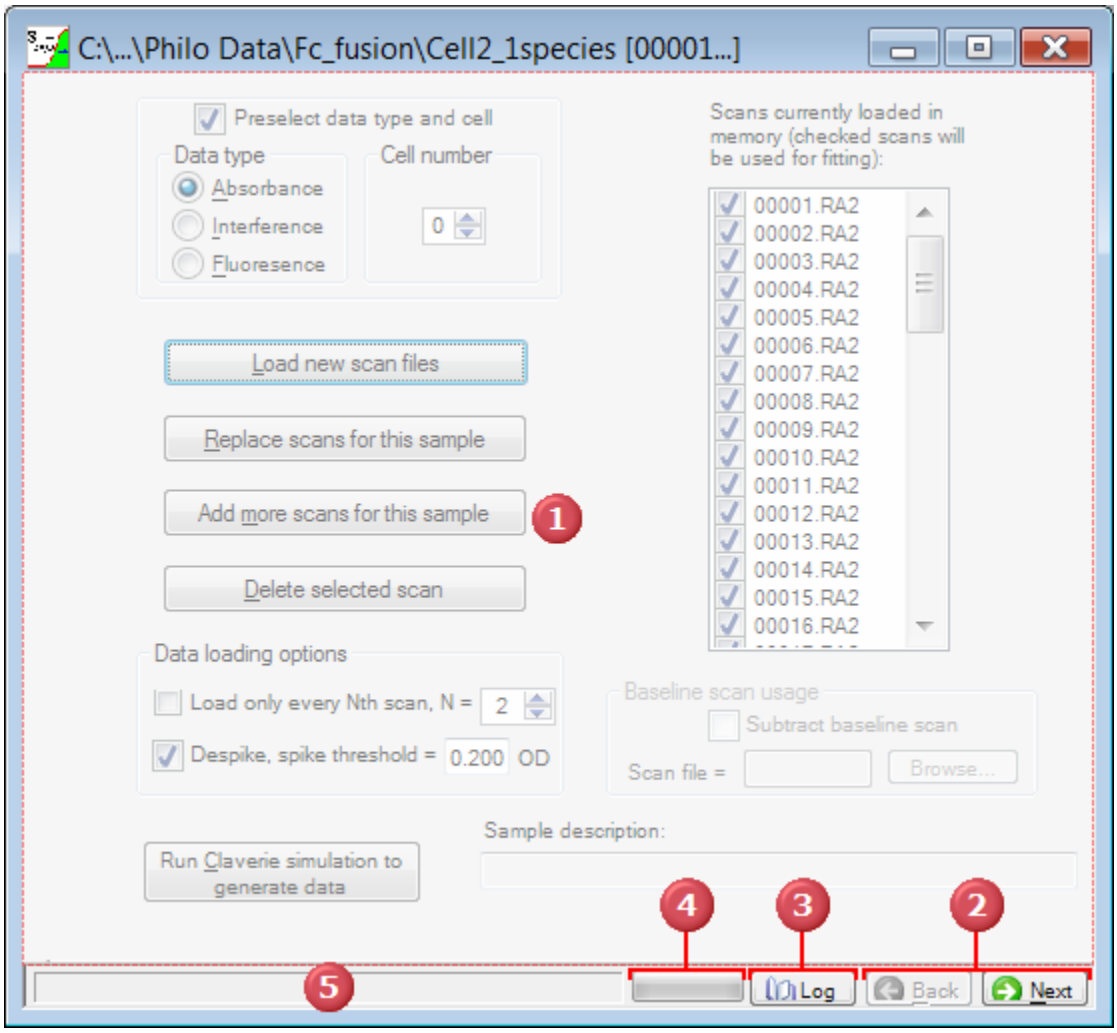
3.1.3 Analysis window

This program allows multiple data analyses to be open at one time. For example, those might be:

1. multiple cells from one velocity run
2. data from multiple velocity runs
3. different fitting models for the same data (same run, cell, and scans), for example models with different numbers of species

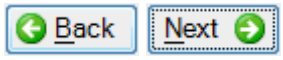
Each of those analyses has an independent analysis window within the main program screen area. Each analysis window shows up to 7 different pages of controls and graphs, depending on the steps within the analysis that have been completed and/or which results you wish to see. Each analysis window can also display its corresponding **Log page (Section 3.1.3.2)**.

1 Those different pages occupy the area that is outlined and washed-out within the image below. The washed-out area shows the first page, the **Load raw scans page (Section 3.2.1)**, a page used for loading the scans to be analyzed.



Moving from step to step

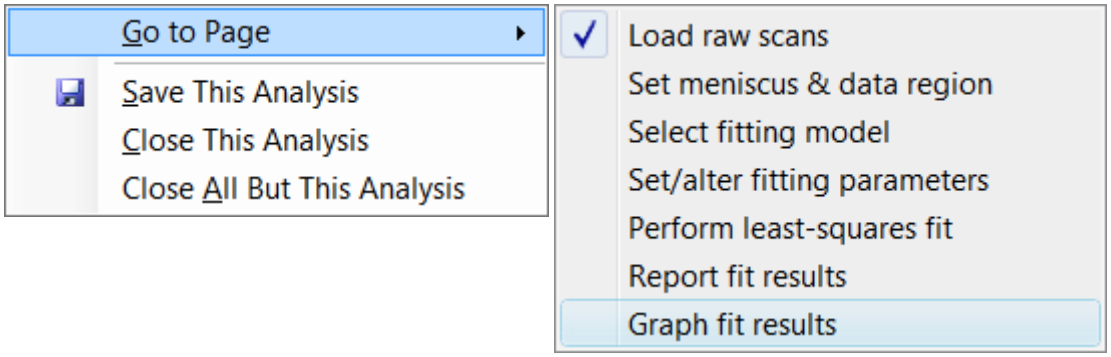
Movement among the steps (pages) within one analysis window can be accomplished using either

2 A) the **Next button and Back button (Section 3.1.3.1)** 

at the lower right,

B) the **Navigation tree (Section 3.1.4)**, or

C) by right-clicking with the mouse on the analysis window (in between the buttons and other controls) and using the **Go to page** pop-up menu shown below. This context menu also has commands for saving this analysis to disk, closing this analysis window, or closing all analysis windows except this one.



Additional tools along window bottom

3 Along the bottom of each window is a panel containing the Log button to display the **Log page (Section 3.1.3.2)**.

4 a **progress bar** that indicates the fractional completion of data loading or lengthy calculations, and

5 a **User Prompt area (Section 3.1.3.3)** that displays instructions or other information. This panel is part of every analysis page.

The panel containing the **Next/Back** and **Log buttons**, progress bar, and user prompt can optionally be moved to the top of the analysis window. Right-clicking the mouse within this panel will bring up a context menu that will toggle this panel between top and bottom locations.

Switching between analysis windows

When multiple analysis windows are open you can switch to a different analysis window by either clicking on that window with the mouse, using the **Navigation tree (Section 3.1.4)**, or using the **Windows menu (Section 3.1.1.5)**.

About window names

New analysis windows are initially named 'Analysis1', 'Analysis2', etc. The current name for each window is shown at its top in its Title Bar area, and also in the corresponding root node within the Navigation tree. Once an analysis has been saved to disk (see **Save Analysis As... (Section 3.1.1.1.3)**) then that path and filename are displayed in the Title Bar. As in the example shown above, when the full path and file name are too long to display, the path is abbreviated.

After scans have been loaded and selected for analysis, the names of the first and last scan that are being used are also shown in that window's Title Bar.

3.1.3.1 Next and Back buttons

The Next and Back **buttons** move forward or back one analysis step (page) at a time.

If the **Next button** is currently disabled this indicates the program is waiting for you to complete some operation on this page that is required before moving to the next step (loading scans, setting the meniscus position, completing a fit, *etc.*). The **Back button** will always be disabled on the first page (**Load raw scans page (Section 3.2.1)**).

3.1.3.2 Log page

As you perform an analysis SVEDBERG creates a Log to record exactly what you have done, and when it was done. This Log can be displayed at any time using the Log button along the bottom of the **Analysis window (Section 3.1.3)**.

The example above shows the first portions of the Log generated during the analysis of some absorbance scans. As you see, each of the steps taken, analyst inputs, and options selected is documented. When a fit is completed, the full report from that fit (including a summary graph for the fit) is placed into the Log.

- **The full Log from the above analysis is contained in this drop-down section as an example.**

Type your Drop Down Section text here.

The Log is a permanent record that is stored along with all the other fit information during the **Save analysis (Section 3.1.1.1.2)** procedure.

If your IT system provides user authentication (the user's identity is validated during a log-in procedure) then the analyst's name (as given by the authentication system) is also added into the log. The user name is added whenever a new analysis is started, or when a previously saved analysis is opened. Thus if user A does the initial analysis, and later user B re-opens the analysis file and re-analyzes the data in a different manner, the log will show who did what (and when).

Buttons and other controls on the Log page

The **Print log** button along the bottom of the log page allows the entire Log, or a selected portion of it, to be printed (to the printer you select from a dialog box).

The **Copy to clipboard** button will place the entire Log, or a selected portion of it, onto the Clipboard in rich text (RTF) format so that it can be pasted into a word processor or electronic notebook software.

- If you want to print or copy only a portion of the Log, first highlight (select) the desired portion by dragging over it with the mouse or using the keyboard movement keys while the SHIFT key is pressed, and then press either the Print or Copy buttons. Copying of the selected portion can also be done using the standard Windows CTRL-C shortcut key combination.

The **Magnification up/down spinner** allows you to increase or decrease the size of the text and images being displayed within the rich text box that displays the Log. This is sometimes useful because:

1. The rich text box sometimes fails to detect that the images are larger than its width and therefore it should display a scroll bar along the bottom.
2. The size of images imbedded in rich text is absolute, which can cause problems when the analysis is saved on one computer and then opened on a computer with a different display resolution.

The **Return** button simply hides the Log page and brings the current analysis page back to the foreground.

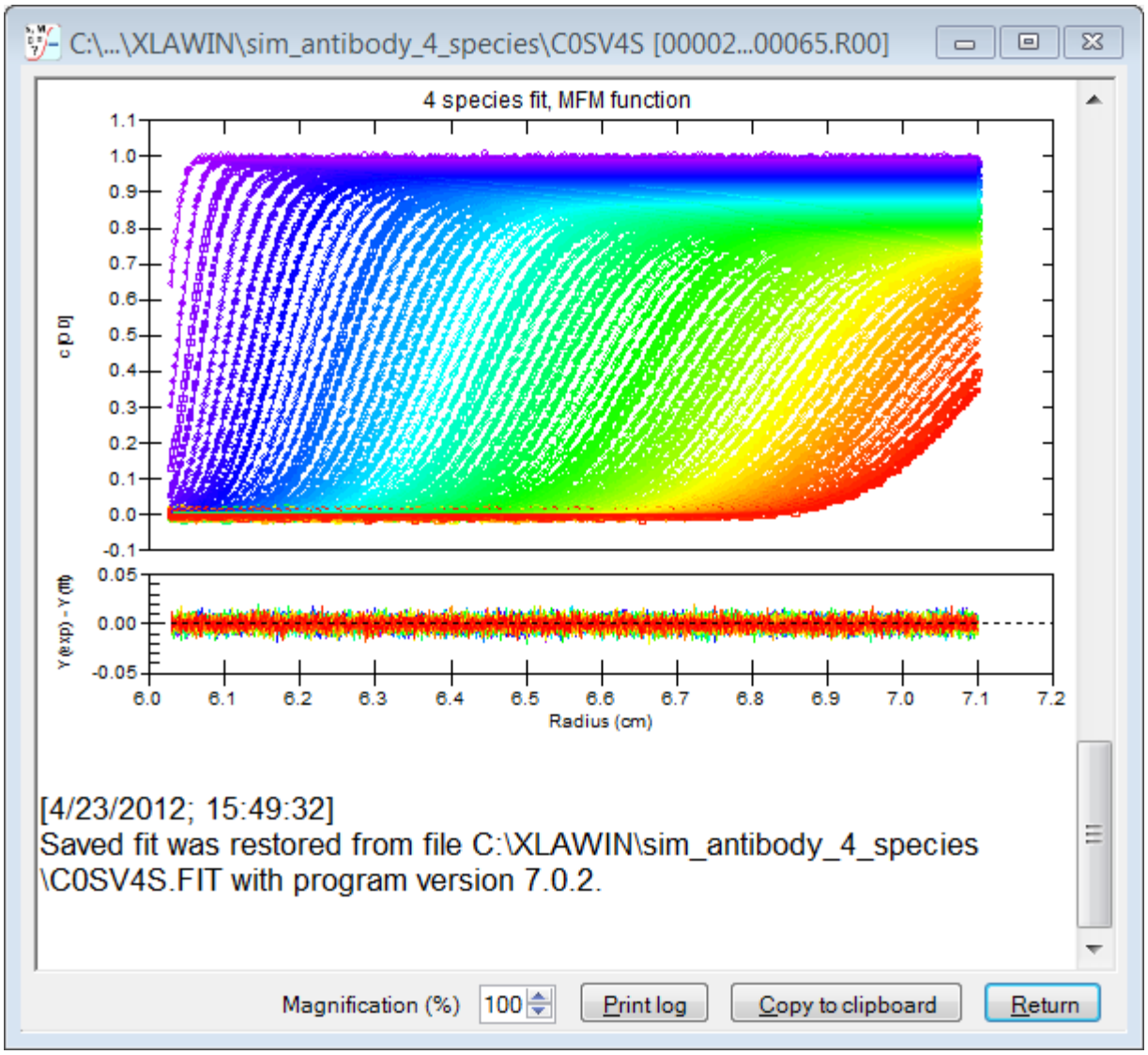
ding graphs to the Log

Typically the various graphs generated during intermediate steps of generating the *g(s*)* distribution, as well as the graphs that can be generated and displayed on the **Graph fit results page (Section 3.2.7)** or by the **Graph bootstrap or Monte Carlo results ({8AE03063-A641-4F0A-B1A1-996EC0C8F64E})** dialog, are considered temporary results and are not placed into the Log.

However if for example one of those graphs illustrates why further refinement of a fit is needed, it may be appropriate to document that finding in the Log before proceeding with such refinements. Therefore buttons are provided on both the **Graph fit results page (Section 3.2.7)** and the **Graph bootstrap or Monte Carlo results ({8AE03063-A641-4F0A-B1A1-996EC0C8F64E})** to add those graphs to the Log when it is desirable to permanently document them.

Loading saved fits lacking a Log

When a fit saved using an earlier program version which lacked the **Log page** feature is loaded a new Log will be created describing the old fit, listing the various parameters such as the fitting region, the names of the scan files, etc. An example of such a case is shown below.



The file name and full path of the restored fit, and the program version that created the new Log, are also entered into the Log (these are not visible in the image above).

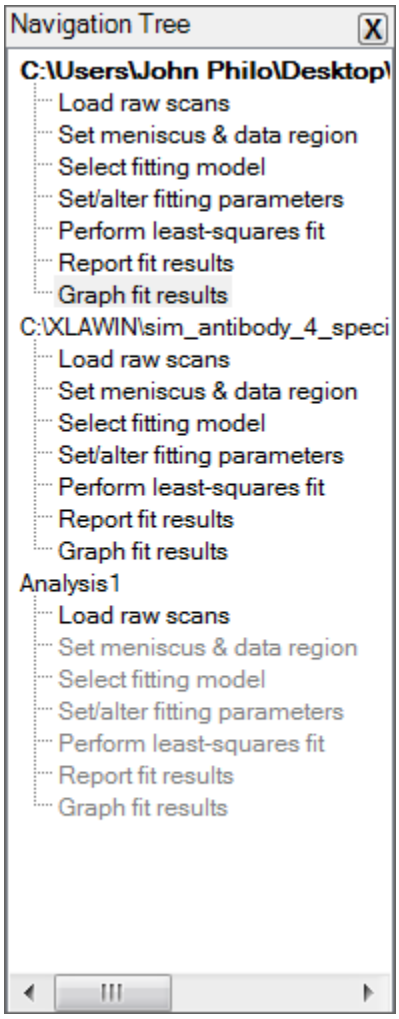
3.1.3.3 User prompt area

The **User prompt area** provides textual prompts for certain actions and displays a position read-out when the user is specifying positions on graphs.

Those prompts will automatically be cleared after being displayed for 5 seconds.

3.1.4 Navigation tree

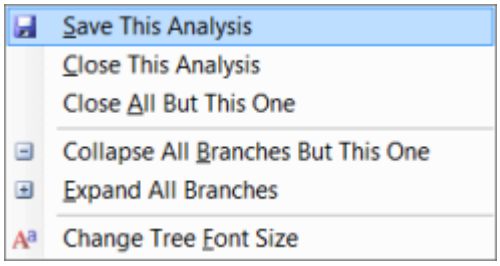
The **Navigation Tree** is a heirarchical tree similar to that used in Windows Explorer and other programs. It provides a root node (branch) for each analysis window, and sub-nodes for the individual steps within an analysis.



Any open **analysis window (Section 3.1.3)** can be selected as the active analysis, and any available step within that analysis can be selected, by clicking on the corresponding node within the tree. Steps which are not yet available are shown as gray (inactive).

The Navigation Tree can optionally be closed to provide more room for analysis windows. Its visibility can be controlled using an entry in the **View menu (Section 3.1.1.3)**. When the width of the Navigation Tree panel is altered by clicking the mouse at the edge and dragging, that width will be remembered the next time the program is started.

Navigation tree context menu


Save This Analysis
Close This Analysis
Close All But This One
Collapse All Branches But This One
Expand All Branches
Change Tree Font Size

Right-clicking on any branch of the tree with the mouse will bring up the context menu shown above. This allows you to:

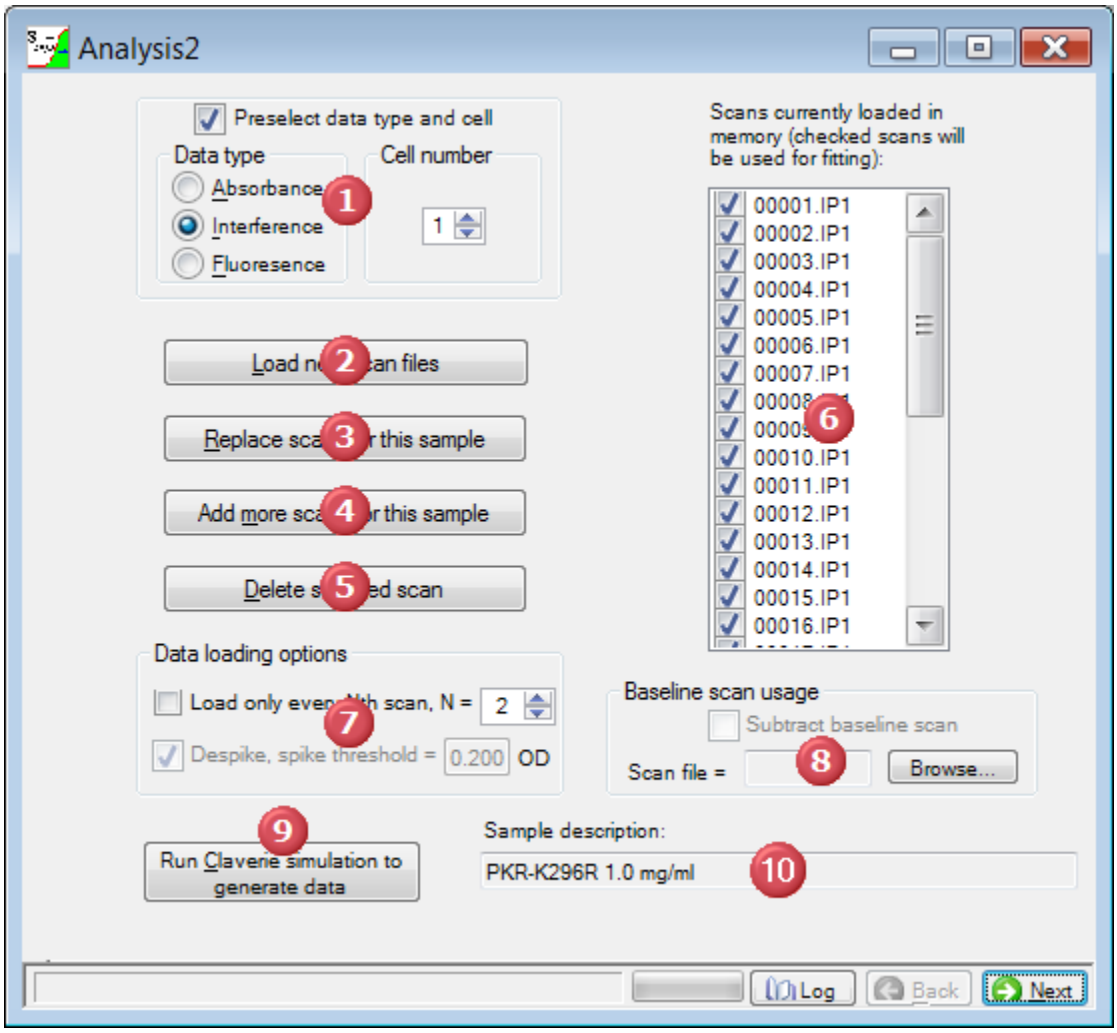
1. save the analysis to disk
2. close this analysis window
3. close all analysis windows except this one
4. collapse all branches of the tree except the one that was clicked
5. expand any collapsed branches in all analyses
6. alter the font size used to display the tree to either make the tree easier to read or to make it more compact

3.2 Analysis pages

3.2.1 Load raw scans page

This page allows you to load data from scan files, and is the starting point for any new analysis. Each control on the image below is a clickable *hotspot* that will lead to a separate page with further information.

Note that the **Next** button will not be active unless at least 1 scan has been loaded.

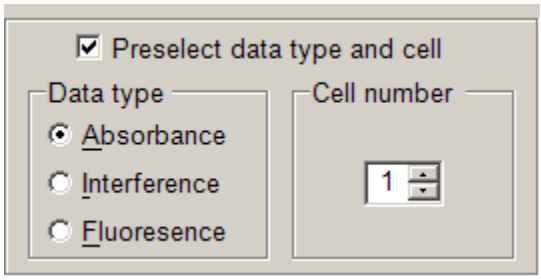


Controls on this page:

1. **Preselect data type and cell check box (Section 3.2.1.1)**
2. **Load new scan files button (Section 3.2.1.2)**
3. **Replace scans for this sample button (Section 3.2.1.3)**
4. **Add more scans for this sample button (Section 3.2.1.4)**
5. **Delete selected scan button (Section 3.2.1.5)**
6. **Scans currently loaded in memory list box (Section 3.2.1.6)**
7. **Data loading options (Section 3.2.1.7)**
8. **Load raw scans - Baseline scan usage (Section 3.2.1.8)**
9. **Run Claverie simulation to generate data button (Section 3.2.1.10)**

10 Sample description (Section 3.2.1.9)

3.2.1.1 Load raw scans - Preselect data type and cell check box



When this box is checked then the **Data type** radio buttons and **Cell number** up-down spinner controls within this group will control the data type and cell number of the scans that are initially displayed in the File Open dialogs.

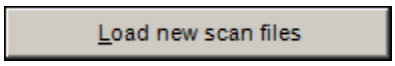
Using this option makes it easier to find the data you are seeking among the hundreds of scan files that may be present. If you change your mind after you are already working within the File Open dialog you can still choose different data types or cells by making a different choice in the 'Files of type' drop-down list near the bottom of that dialog box. It also allows you to quickly select all the scans for a given cell and data type for loading during the File Open dialog by first clicking any one file and then pressing the CTRL-A 'select all' shortcut.

Note that the 'Absorbance' choice also includes **radial intensity** scans, which can be converted to true absorbance or pseudo-absorbance (for when you are using two samples per cell). See **About using radial intensity scans (Section 6.2)**.

The cell selection spinner allows cell number 0 because SEDFIT creates cell 0 files when it saves simulated data. It also allows cell 8 for those who replace the counterbalance with an additional cell.

The preselect mode is on by default when the program is installed. You can turn off this default on the General Options tab of the **Options dialog (Section 3.1.1.4.3)** .

3.2.1.2 Load raw scans - Load new scan files button



This button is used to initially load scan files into a new analysis window, or to replace all the data that was previously loaded into that window. It will bring up a standard Windows file open dialog which allows you to select and load multiple scan files. Because SEDFIT users are used to using the keyboard shortcut CTRL-L to load scan files, in SVEDBERG that shortcut works to press this button.

Tip 1: Generally you will want to load scans covering the entire range of boundary movement, *i.e.* all or nearly all the scans. If you know your latest scans go long after the cell is empty, there is no point in loading those, and you also may want to skip the first ~5-10% of the run where the meniscus region is barely cleared.

Tip 2: If you have hundreds of interference or fluorescence scans, at least initially to save time you may want to load only every second or third scan by using the **Load every Nth scan (Section 3.2.1.7)**

option.

Tip 3: If you have already analyzed this experiment using SEDFIT, you can load the same scans by specifying the SEDFIT list file (list.ra1, list.ip1, *etc.*) in the File Open dialog. After loading the files specified within the list file, if a corresponding SEDFIT parameters file is also found, the program will ask whether you also wish to load the meniscus position, cell base position, and data fitting region used in your SEDFIT fitting.

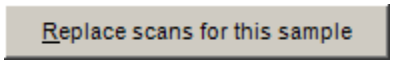
Tip 4: SVEDBERG recognizes that SEDFIT creates temporary files with file extensions that suggest they are scan files (files with names such as ~tmppars.ra1, list.ra1, and sdist.ra1), and will ignore them if you happen to select them as scans to be loaded. Together with the **Preselect cell and data type option (Section 3.2.1.1)** this allows you to quickly and easily load all available scans by clicking any file and using the Windows shortcut key CTRL-A to select all of them.

If the **Preselect data type and cell (Section 3.2.1.1)** option is on (checked) then the File Open dialog will initially show only those scan files that match that data type (absorbance, interference, or fluorescence) and cell number.

You can always change the data type and cell number that are being shown, or display all scan files, by picking appropriate choices from the 'Files of type' drop-down list near the bottom of the File Open dialog box.

The initial scan folder that will be displayed depends on settings within the File Locations tab of the **Options dialog (Section 3.1.1.4.3)**. You may choose either to always start in a particular folder (for example the folder where your instrument deposits new data), or to have the program remember the location of the last data that were loaded and to always start there.

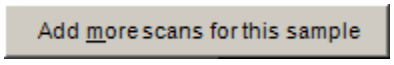
3.2.1.3 Load raw scans - Replace scans for this sample button



This button is used to load a different set of scans for a particular sample, retaining sample-dependent settings such as the meniscus location. It is the retention of those settings that distinguishes this operation from using the **Load New Scan Files button (Section 3.2.1.2)** to replace the scans.

This button is only active after scans have been loaded using the **Load New Scan Files button (Section 3.2.1.2)**.

3.2.1.4 Load raw scans - Add more scans for this sample button

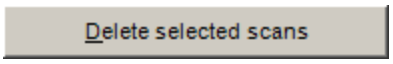


This button allows you to load additional scans while retaining sample-dependent settings such as the meniscus position.

The added scans can be from times either earlier or later in the run than those already loaded. The scan files are sorted based on the run times at the end of this operation, and any duplicate scans are removed.

This button is only active after scans have been loaded using the **Load New Scan Files button (Section 3.2.1.2)**.

3.2.1.5 Load raw scans - Delete selected scans button

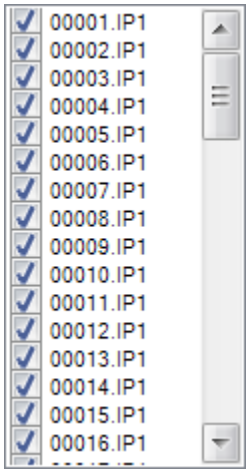


This button will delete the scan which is currently highlighted in the **Scans currently loaded in memory list box (Section 3.2.1.6)**. (If nothing is highlighted a beep is emitted but nothing else happens.)

This command is useful if you need to remove an entire scan where the data is bad (for example an interference "*rogue scan*" that bears no resemblance to the others). Alternatively, a bad scan can be removed from use in this analysis (but not totally removed from memory) by un-checking the check box beside that scan's file name in the **Scans currently loaded in memory list box (Section 3.2.1.6)**.

3.2.1.6 Load raw scans - Scans currently loaded in memory list box

This list box shows the scan file names for all scans currently in memory for this analysis window. It is possible to temporarily exclude scans from being used in the data analysis, but still keep them in memory, by clicking on the check box next to the scan name to un-check it. For example, this can be used to exclude a *rogue scan*.

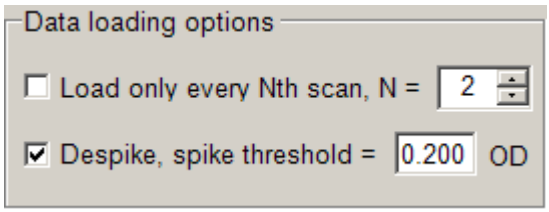


If you right-click on any scan name a pop-up context menu will appear with options to:

1. display the scan header information (comment, wavelength, run time, total number of data points, etc.)
2. edit the scan file using NotePad

If you choose to edit the scan file in NotePad the program will minimize itself and launch NotePad with that scan loaded. The program then waits for NotePad to close. When NotePad closes, the program detects whether the file was altered, and if so it automatically re-loads the altered data. For other data editing options see **How to remove bad data points from scan files (Section 4.4)**.

3.2.1.7 Load raw scans - Data loading options



This group of controls specifies optional parameters that alter how scan files are loaded.

If the **Load every Nth scan** check box is activated (checked) then only every Nth scan among the scans selected within the File Open dialog will be loaded when new scans are read. This applies to the **Load New Scan Files button (Section 3.2.1.2)**, **Replace Scans for this Sample button (Section 3.2.1.3)**, and the **(Section 3.2.1.4) Add More Scans for this Sample button (Section 3.2.1.4)**. The value of **N** (default = 2) can be altered using either the up-down spinner or by directly typing a number within its text box.

Normally hundreds of scans can be loaded and processed if desired (limited only by available memory). Thus the use of this option is almost never really required. This option is primarily intended to speed up data processing for situations where (1) a very large number of interference scans have been acquired and there is very little boundary movement between successive scans, and (2) the signal/noise of those scans is sufficiently high that a large number of scans is not required.

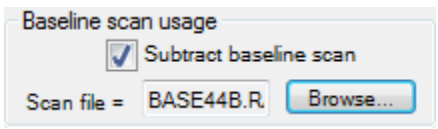
The **Despike options** are only active for absorbance scans. By default despiking is always on. The purpose of despiking is to remove the single-point spikes that sometimes occur in absorbance scans (often due to lamp miss-firing). These spikes are visible as single-point spikes within the dc/dt curve from one pair of scans (and a corresponding smaller spike in the average dc/dt curve), as displayed on the 'Select scans to analyze' page.

The despiking routine defines spikes as data points which differ from the preceding and following points by more than a threshold value, and removes those data points (from the data being processed, not in the original scan file). This **spike threshold** defaults to 0.2 OD, but can be altered using the corresponding text box. The default value works well in most situations, but may fail to detect some small spikes. The danger of setting the threshold too low is that false spikes may be detected (for example normal random noise in the data, or the top of the meniscus region). All of the original raw data is always retained in memory (and will be stored by the **Save Analysis As... (Section 3.1.1.1.3)** command), so you can alter this setting as desired and reprocess the data until you are satisfied with the scans displayed on the **Set meniscus & data region page (Section 3.2.2)**.

Note that the despiking routine ignores a small region around the meniscus so the meniscus spike is not affected.

If you want to see all the bad data in the graph on the **Set menisus & data region page (Section 3.2.2)** (for example if you want to edit the raw scans to eliminate bad points) then you will want to turn despiking off.

3.2.1.8 Baseline scan usage



This group of controls allows you to load a baseline scan that will be subtracted from each and every scan used in the analysis. The **Browse...** button will bring up a standard File Open dialog box to designate which baseline scan to load, and once a scan is loaded its use can be turned on or off with the check box above.

To remove an existing baseline file from memory push the **Browse...** button and then Cancel the File Open dialog.

What exactly is a baseline scan and why would I use one?

Sedimentation velocity scans often have baseline offsets (a constant signal at all radii when the concentration is zero, also known as "radially-invariant noise" or RIN). They also typically have some baseline distortions and/or "window noise" from scratches, dirt, or fingerprints that cause signals that vary at different radii across the cell but which are constant from one scan to the next (also known as "time-invariant noise", "time-independent noise", or TIN). One way (and often the best way) to eliminate these systematic artifacts is to subtract an appropriate experimental baseline scan from each data set.

Another approach is to remove this systematic noise arithmetically using scan differencing or 'delta-*c* mode' (see **Select fitting model - Data options (Section 3.2.3.4)**). That approach often works well, but has these disadvantages:

1. it mixes together information from scans both early and late in the run, making it difficult to interpret overlays of the fitted data and the fit and residual plots
2. it slightly reduces the signal/noise ratio (theoretically by the square root of 2) because the photometric noise from 2 independent scans contributes to each delta-c data set, and it reduces the total number of fitted data points by a factor of 2.

A third approach is to introduce extra fitting parameters and actually fit an unknown baseline (time-invariant noise, TIN) as part of the overall fitting process (see **Select fitting model - Baseline fitting options (Section 3.2.3.5)**). However a significant drawback of this approach is that the computed baseline is model dependent. If the fitting model is correct, and if you obtain a good fit (no systematic deviations) then this fitting procedure will give the correct baseline, but when the model is wrong the fitted baseline tries to compensate for the model's inadequacies by distorting the baseline, which can potentially even lead to bad outcomes such as false species.

Note that versions of SVEDBERG prior to version 7.0.0 allowed the possibility of multiple baseline scans (different baselines for different scans). That feature is no longer implemented since by definition a baseline is constant in time. When reading saved analyses from earlier versions the program will now apply the first baseline scan it finds in the old saved fit to all the scans.

How can I obtain a baseline scan?

Baseline scans can be obtained, for example, after a run by rinsing out the sample sector several times, refilling it with the same solvent used in the reference channel, and then recording a scan at the same rotor speed used for that sample. Two drawbacks of that approach are:

1. It is difficult to re-fill the sample channel to exactly the same column height. For absorbance scans it is usually sufficient to fill to a height greater than or equal to that in the actual experiment, but for interference scans the baseline distortions are sensitive to the exact column height.
2. In removing and replacing the cell in the rotor you may not get exactly equivalent positioning.

An alternative approach that works for relatively high molecular weight solutes, or ones that tend to precipitate as they accumulate at the cell base, is to simply continue the run until all of the solute is pelleted and then record your baseline scan. Further discussion of obtaining and using baseline scans can be found in **Reference 1 (Section 7.1)**.

It is advantageous to use signal averaging to reduce the photometric noise in the baseline scan. The recommended approach is to repeat the baseline scan many times, using settings identical to those used when acquiring the velocity data for that sample, and then average together the scan files. One utility program that can do this is **XLGRAPH** (which can be downloaded for free from the same page where this program is downloaded). For absorbance scans it is also possible to tell the GUI to average multiple flashes, but this may introduce shifts in the radius values relative to those recorded during the velocity run, which would degrade

subtraction of sharp features in the baseline profile such as scratches on the windows.

How is the subtraction handled if the experimental scan radial positions or range don't match those of the baseline scan?

For absorbance scans the data points in the baseline scan will often not fall at exactly the same radii as in the data points in the main scan. When this happens a linear interpolation of the baseline scan is done, except when the main scan covers a larger radial region than the baseline scan. In this latter circumstance, no extrapolation of the baseline is done, and the values for the first or last data point in the baseline scan are simply applied to any region where baseline data do not exist.

3.2.1.9 Load raw scans - Sample description

This text display area shows the sample description from the comment line of the last scan that was loaded (the information that was entered for that cell during data acquisition), or the description of the last Claverie simulation if simulated data are currently loaded.

3.2.1.10 Load raw scans - Run Claverie simulation to generate data button

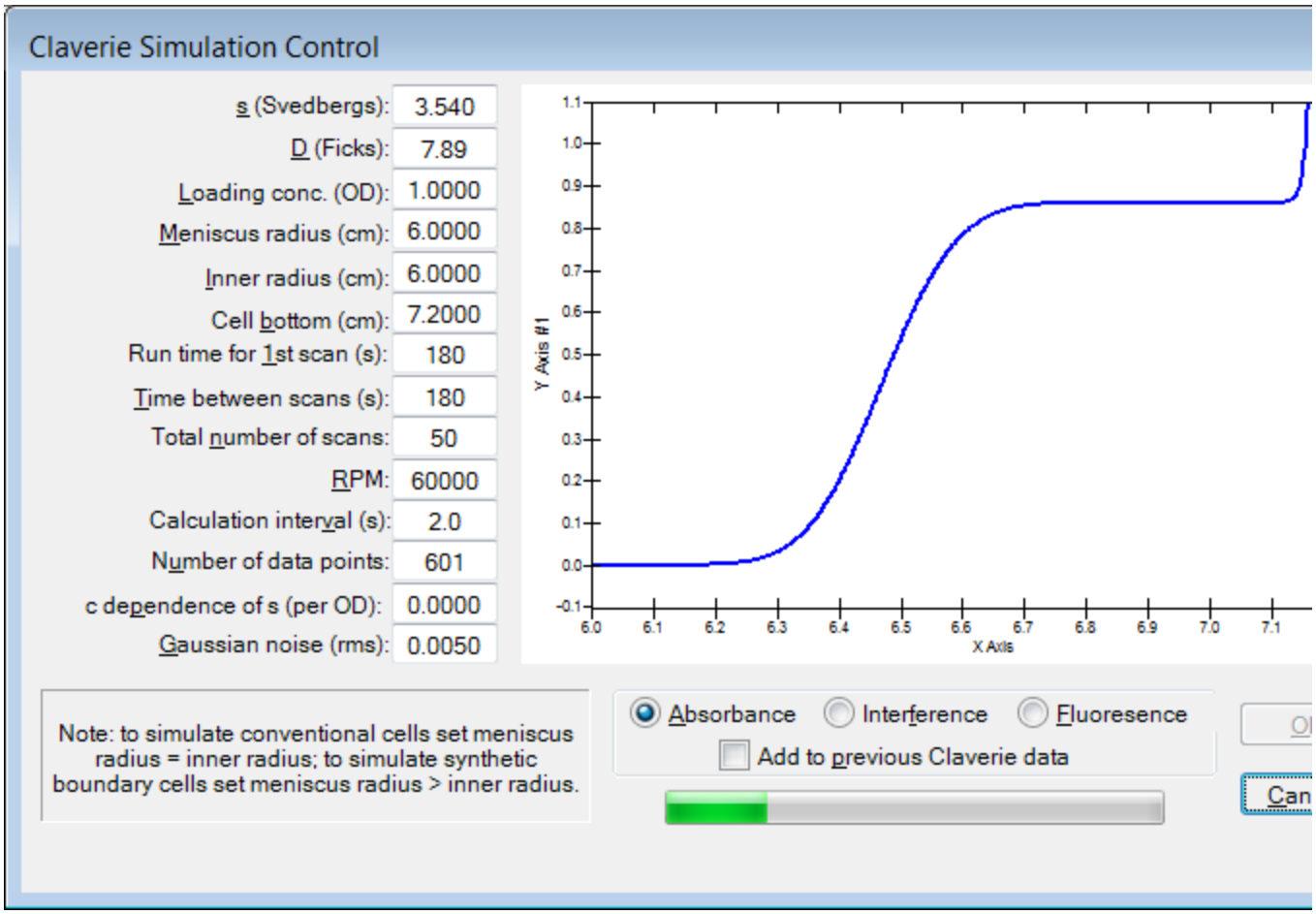
A Claverie simulation uses finite-element methods to solve the Lamm equation. This approach can be give highly accurate simulations of velocity experiments, but is computationally intensive.

The purpose of providing this simulator is to provide a tool for "what if" explorations and tests of the accuracy of the fits. Some examples of questions that can be approached this way are:

1. Is the difference between the mass value returned by fitting g(s*) and an expected result really significant, or just due to inaccuracy of the method?
2. To what extent will adding more scans to your analysis cause broadening of the peaks and therefore inaccurate D and M values?
3. What would the data and residuals look like if my sample contained a small amount of a second species, but was analyzed as a single species?

This Claverie simulator (code kindly provided by David Cox and Walter Stafford) allows you to generate accurate data for these scenarios, which you can then use for fitting, comparing the fitting results to the known correct answers. It also allows for concentration-dependent sedimentation coefficients (something not dealt with by the fitting functions) and thus permits you to see how concentration dependence will affect the results. It does not, however, include concentration dependence of diffusion coefficients.

The dialog box used to control and run the simulations is shown below as it appears while a simulation run is in progress. This run corresponds to the default parameters, a quick (moderate accuracy) simulation of ovalbumin at 60K rpm.


Claverie Simulation Control
s (Svedbergs): 3.540
D (Ficks): 7.89
Loading conc. (OD): 1.0000
Meniscus radius (cm): 6.0000
Inner radius (cm): 6.0000
Cell bottom (cm): 7.2000
Run time for 1st scan (s): 180
Time between scans (s): 180
Total number of scans: 50
RPM: 60000
Calculation interval (s): 2.0
Number of data points: 601
c dependence of s (per OD): 0.0000
Gaussian noise (rms): 0.0050
Y Axis #1
X Axis
Note: to simulate conventional cells set meniscus radius = inner radius; to simulate synthetic boundary cells set meniscus radius > inner radius.
Absorbance
Interference
Fluoresence
Add to previous Claverie data

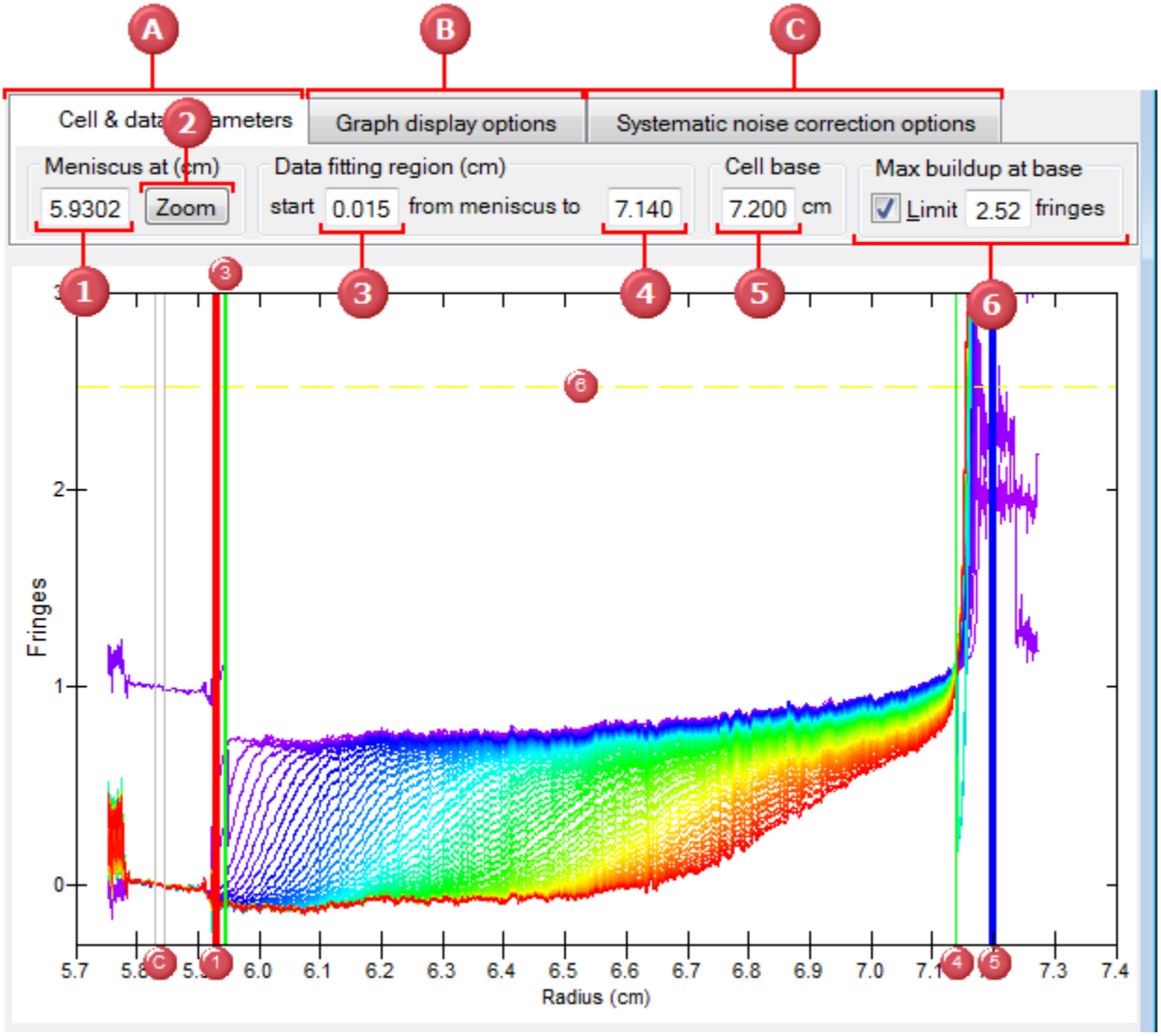
For instructions on how to use this simulator see **Tutorial: Simulating experiments with the Claverie simulator (Section 2.4)**.

3.2.2 Set meniscus & data region page

This page allows you to view a graph of the scan data, set the positions of the meniscus and cell base, and set the radial limits defining the data to be fitted. A number of less commonly used controls and options are organized onto the other three tabs of a 4-tab control (marked B-D in the image below). The "A" tab shown below, the **Cell & data parameters tab**, is always displayed first whenever this analysis page is first displayed.

When this page is first displayed with new data, the **meniscus wizard (Section 6.4)** automatically attempts to set the meniscus position, but manual adjustment is always needed for fluorescence scans, and may be needed for interference or absorbance scans.

Each control on the image below, as well as the vertical position markers on the graph, are clickable *hotspots* that will lead to further information, or just use the numbered list below the image.



Controls on this page:

Tab Pages:

A The **Cell & data parameters tab** displays controls for setting cell parameters (meniscus and cell base positions) and the upper and lower bounds for the regions to be fitted.

B The **Graph display options tab (Section 3.2.2.1)** brings up controls that allow showing only every N-th scan (to speed up the display or to make a better figure for a presentation or publication), the alternate display of a 3-D version of the graph, and a tool to identify individual scans and data points within the graph (and optionally edit individual points when needed).

C The **Systematic noise corrections options tab (Section 3.2.2.2)** brings up controls related to the algebraic removal of *fringe jumps* and/or *jitter* (RIN) from interference scans or radially-invariant noise from *pseudo-absorbance scans*. This tab is not displayed for standard absorbance scans or for fluorescence scans.

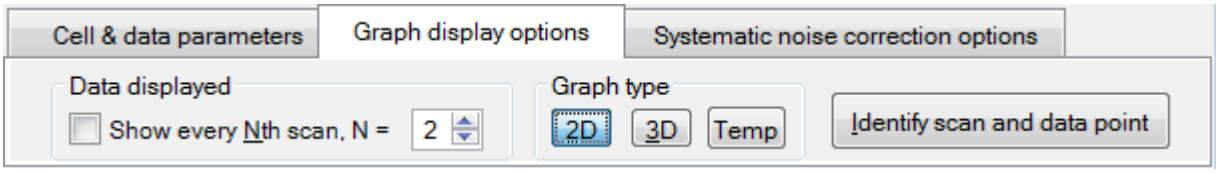
Controls displayed for the Cell & data parameters tab:

1. **Meniscus position (red marker on graph) (Section 3.2.2.3)**
2. **Zoom meniscus button (Section 3.2.2.6)**
3. **Good data begin X cm from meniscus (left green marker on graph) (Section 3.2.2.4)**
4. **Good data end at radius = (right green marker on graph) (Section 3.2.2.5)**
5. **Set meniscus & data region - Cell base position (Section 3.2.2.7)**
6. **Set meniscus & data region - Limit maximum buildup at cell base (Section 3.2.2.8)**

Note 1: By default when large numbers of scans are loaded in order to speed up the graph display and graph interactions the program will limit the graph to < 100 scans by displaying only every second, third, *etc.* scan. It is always possible to display all the scans by switching to the **Graph display options tab (Section 3.2.2.1)** and turning off the **Show every Nth scan** option. This default behavior can also be turned off by a choice on the **Other options tab** of the **Options dialog (Section 3.1.1.4.3)**.

Note 2: to keep the graph legend from occupying too large a fraction of the graph the legend is suppressed when more than 40 scans are displayed (as shown above). When more than 40 scans are shown you can force display of a legend by right-clicking on the graph and selecting a legend position from within the Legend Style submenu. The legend text size is reduced when there are 20 or more scans shown.

3.2.2.1 Graph display options tab



The **Graph display options tab** of the **Set meniscus & data region page (Section 3.2.2)** shows controls that allow showing only every Nth scan (to speed up the display or to make a better figure for a presentation or publication), the alternate display of a 3-D version of the graph, and a tool to identify individual scans and data points within the graph (and optionally edit individual points when needed).

Show every Nth scan

When this box is checked only every N-th one of the total scans to be used in this analysis are displayed. When the number of scans is large, and especially for interference scans, this option can greatly increase the speed of graph operations such as dragging position markers. It is also often useful when preparing raw data graphs for use in presentations or publications because showing all the scans produces too much overlap.

The value of N can be changed using the up-down spinner or by directly entering a number into the text box area next to the up-down spinner.

By default in order to speed up the graph display and graph interactions when large numbers of scans are loaded the program will limit the graph to < 100 scans by automatically turning on the **Show every Nth scan** when needed, and also by adjusting N as needed to keep the total displayed scans below 100.

It is always possible to display all the scans by switching to the **Graph display options tab (Section 3.2.2.1)** and turning off the **Show every Nth scan** option. This default behavior can also be turned off by a choice on the **Other options tab** of the **Options dialog (Section 3.1.1.4.3)**.

Graph type options

By default the scans are shown as a conventional 2-dimensional line graph, and hence by default the **2D button** is depressed (on).

Pressing the **3D button** will bring up a 3-dimensional "waterfall" graph of the scans, with the color of the filled area below successive scans varied to produce a rainbow spectrum. In this graph only the portions of the scans within the analysis region (between the green vertical bars on the 2D graph) is shown. The viewing angle for the 3D graph can be rotated using the scroll bars on the right and bottom of the graph; the axes are suppressed during the rotation operation.

- This 3D display really has no function for data analysis, it is simply an alternate method of displaying the data and the motion of the boundaries. The positions of the meniscus and the other region markers are not displayed on this 3D graph.

Pressing the **Temp button** will bring up a line graph of the recorded rotor temperature versus run time.

TIP: Remember that temperature accuracy and stability is critically important for accuracy and reproducibility of the sedimentation coefficients. This temperature plot is a convenient way to check whether your run had good temperature stability.

Identify scan and data point button

This button will cause the 2D graph of the raw data to be displayed as individual data points rather than continuous lines, and each data point becomes a clickable object. When a data point is clicked the corresponding scan file name and *x,y* data values are displayed within a box above the graph area.

The main purpose of this display mode is to help identify *rogue scan*s (see **How to identify and remove rogue scans (Section 4.5)**) or individual bad data points.

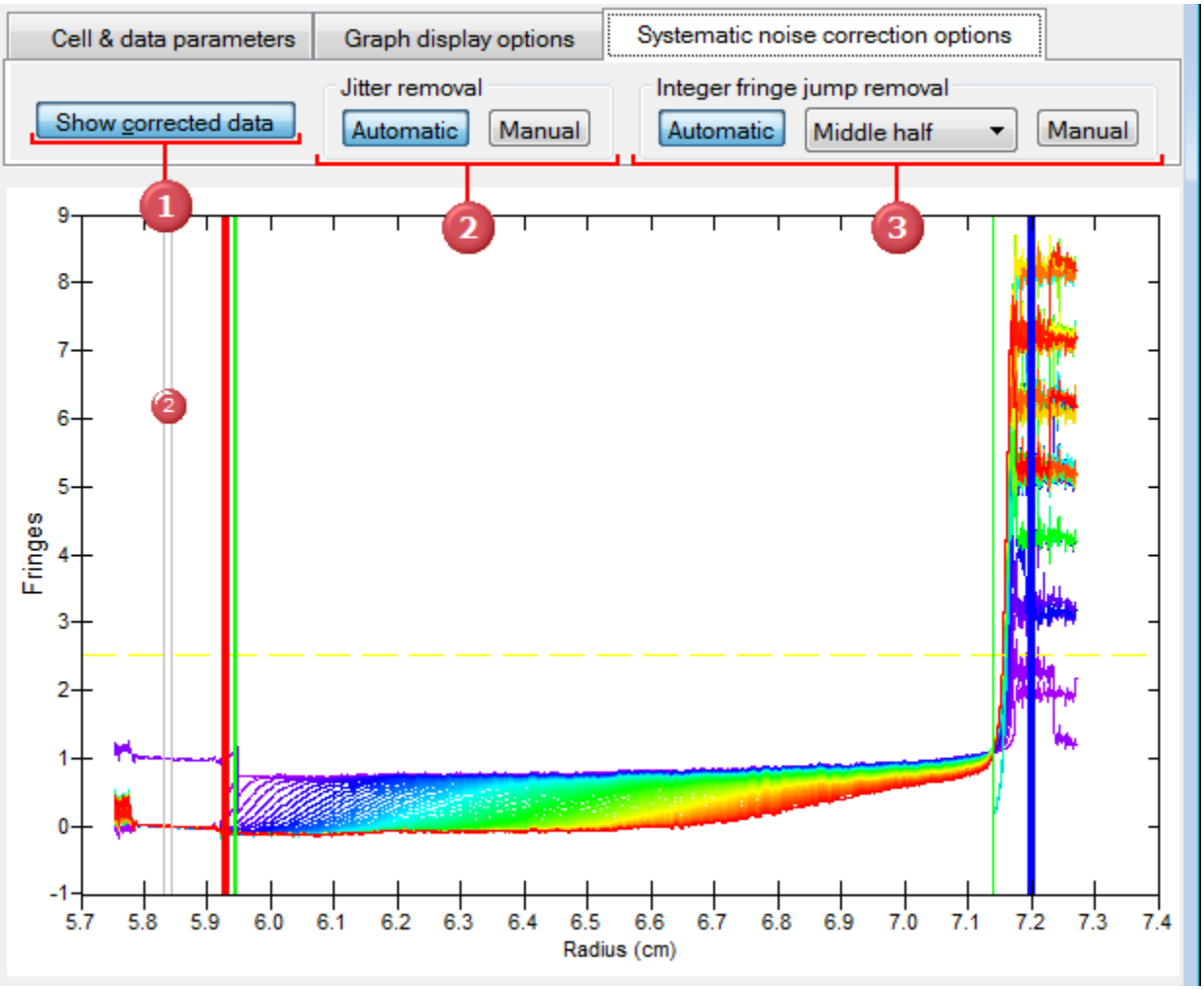
Note that for graphs containing large numbers of scans, especially interference scans, there can be a significant delay (10 sec or more) between clicking the button and the re-display of the graph, and also between clicking a data point and the display of its identity. This is simply because there are tens to hundreds of thousands of clickable regions to be created and then identified.

When the same data point is clicked two times in succession (separate single clicks, not a double-click) the program assumes this may be a bad data point that you want to remove. You will be asked a series of questions about whether you want to remove that point only temporarily (from memory), or to permanently alter the scan file (either reversibly or irreversibly). See **How to remove bad data points from scan files (Section 4.4)** for more details.

3.2.2.2 Systematic noise corrections options tab

The **Systematic noise corrections options tab** of the **Set meniscus & data region page (Section 3.2.2)** shows controls related to the algebraic removal of *fringe jumps* and/or *jitter* (*RIN*) from interference scans. Because *pseudo-absorbance scan*s also can have vertical displacements similar to *jitter* the **Jitter removal** group box is also enabled for *pseudo-absorbance scan*s.

Each control group on the image below, as well as the gray vertical position markers on the graph, are clickable *hotspot*s that will lead to further information, or just use the numbered list below the image.



Controls displayed for the Cell & data parameters tab:

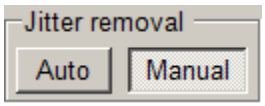
1 **Show corrected data button -** When selected this option displays the scan data after the jitter and/or fringe jumps have been removed, which is the normal program default. This button acts as a toggle switch, with the uncorrected 'raw' scans shown when it is off (up).

2 **Jitter removal options (Section 3.2.2.2.1) -** Automatically or manually set the radial region used to vertically align the data

3 **Fringe jump removal options (Section 3.2.2.2.2) -** Automatically or manually set the radial region used for locating and correcting integer fringe jumps

3.2.2.2.1 Set meniscus & data region - Jitter removal options

Jitter removal options are displayed only on the **Graph display options tab (Section 3.2.2.1)** and active only for interference data and pseudo-absorbance data.



The purpose of *jitter* removal is to correct for small fringe shifts (vertical displacements of small fractions of a fringe) that often occur from one interference scan to another. The correction is done by selecting a region (a range of radii) where the scans are expected to align vertically, and then forcing them to do so. This type of matching is also called "normalization" in the Beckman Coulter software, and "fine adjusting" in some versions of Walter Stafford's SEDANAL program.

A similar phenomenon also seems to arise when absorbance data are collected in pseudo-absorbance mode (two samples per cell). In that case the vertical shifts typically occur in a few discrete jumps per run.

The jitter matching region is generally placed either in the air-air region or a region just to the right of the meniscus. Using a region to the right of the meniscus generally works better. However, this will only work correctly if the meniscus is completely cleared and there is no sedimentation of low-molecular-weight components that will produce signals in this region.

By default when new interference data are loaded the **jitter wizard (Section 6.5)** tries to automatically determine the best matching region, marks that region on the graph using two vertical gray bars to mark the region limits, and then does the correction using that region. The **Auto button** will show as depressed (on) to indicate this automatic correction is active.

- The jitter wizard will be invoked as soon as you set a meniscus position, either by using the Enter key to confirm the program's automatic determination of the meniscus location, or by manually dragging the red meniscus marker.

To manually select the match region:

Two gray vertical bars are displayed on the graph to mark the inner and outer radii for the match region. When the **Manual button** is pressed a prompt is given to drag those gray markers to a new position. Whenever the markers are dragged away from the position set by the jitter wizard, the Manual button will change to the depressed (on) state.

When you drag either gray bar both bars will move together, maintaining the same separation between them.

If you drag the gray markers **while holding the Shift key down** the two gray bars can be moved independently.

Once the Manual button is pushed **this operation must be completed** by dragging a gray marker; until it is completed other controls will be inactive.

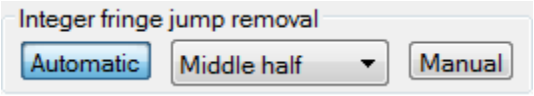
Note that moving the right marker far to the right of the meniscus will affect how much raw data is included in the fit, and thus could affect the ability to accurate define the most slowly-sedimenting species. Only data points which are to the right of both the leftmost green marker and the rightmost gray marker will be used for data analysis. (Since the data in the match region are being forced to be zero, it makes no sense to include them in the analysis.)

To turn off jitter correction:

Jitter correction is off whenever both the **Auto button** and the **Manual button** are up (off). Simply press whichever button is down to turn jitter correction off.

3.2.2.2.2 Set meniscus & data region - Fringe jump removal options

Fringe jump removal options are displayed only on the **Graph display options tab (Section 3.2.2.1)** and active only for interference data.



This group of controls selects the method used to compensate for any *fringe jumps* that occur within a group of interference scans that are being analyzed. The methods differ only in the region of the scan that is used in comparing successive scans for a possible fringe shift.

Four 'automatic' methods are selectable via a drop-down list box. Those are:

1. **Entire scan** compares the scans across the entire sample area to the right of the meniscus. This will work well as long as the amount of sample movement between scans is not too large (moderate concentrations and sedimentation rates), which is generally true for scans appropriate for the dc/dt method.
2. **Near meniscus** confines the comparison to the part of the sample near the meniscus (60 data points). This will work well if all the scans are taken after the meniscus region has cleared.
3. **Middle half** compares the scans in the region from 1/4 to 3/4 of the sample column. This is the default method when the program is initially installed.
4. **Outer eighth** makes the comparison only for the high radius portion. This will work well if all the scans have a well-defined plateau region, provided the scans do not include too much data from the region where sample is accumulating at the bottom of the cell. The outer end of the range used for this comparison is limited by the 'good data end at radius' setting.

A default automatic method will automatically be used whenever this analysis window is first entered after new scans have been loaded.

Whenever an automatic method has been used the **Auto button** within this group will be depressed. If a different method is selected via the drop-down list box while the Auto button is depressed, the new method will be calculated and applied to the data displayed on the graph.

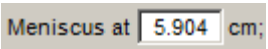
Manual fringe jump removal

The **Manual button** initiates a procedure where the position used for fringe jump compensation is specified by double-clicking on the graph at the desired radial position. The actual region used in the calculation is then a group of 11 points centered around that position. This manual setting will change the 'Auto' button to the off (raised) position.

Once the Manual button is pushed **this operation must be completed** by double-clicking on the graph; until it is complete other controls will be inactive.

All these methods are implemented in the same manner as the corresponding methods in Walter Stafford's original DCDT program. The default method can be altered via the 'Cell & Data Options' tab within the **(Section 3.1.1.4.3)Options dialog (Section 3.1.1.4.3)**.

3.2.2.3 Set meniscus & data region - Meniscus position



This setting defines the radial position of the meniscus. The value can be altered either by directly entering a value in the text box, or by dragging the red vertical line displayed on the 2D graph. The position entered on this analysis page will be automatically entered as the meniscus position fitting parameter (if you choose to fit the meniscus position then this represents the initial guess value).

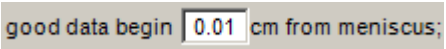
- For absorbance scans, if you manually drag the meniscus marker on the graph you will then be asked whether you want the wizard to re-calculate the mean position of the positive spike (the mean of the actual spike positions for every scan that has been loaded). This is helpful when the scans contain two positive spikes, and the wizard has identified the wrong one as the true meniscus.

Upon first entry to this analysis page after new scans have been loaded the **meniscus wizard (Section 6.4)** will try to locate the meniscus for you, and the *focus* will be moved to the meniscus entry text box. Thus to confirm the value set by the wizard all you need to do is push the **Enter** key on the keyboard.

- If these are interference scans using the **Enter** key will then also trigger the **jitter wizard (Section 6.5)**, an automatic **fringe jump correction (Section 3.2.2.2.2)**, and an update of the graph after these systematic errors are removed. If the wizard has not done a good job and rather than accepting its choice you immediately use the mouse to drag the meniscus marker to a new position, those automatic systematic noise corrections will be postponed until the end of that meniscus drag operation.

You may want to zoom in on the region around the meniscus. You can do this manually (see **How to zoom with the mouse (Section 4.3.4)**) or the Zoom meniscus button can do this for you.

3.2.2.4 Set meniscus & data region - Good data begin X cm from meniscus

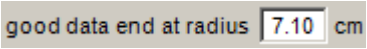


This setting is used to exclude the region near the meniscus where the data are distorted by optical artifacts. As you adjust the meniscus position the left hand green marker moves along with it to maintain a constant separation.

This value can be altered either by directly entering a value in the text box, or by dragging the left-hand green vertical line displayed on the 2D graph.

Obviously if you find that you often need to adjust this value then you may need to revise the default value via the **Cell & Data Options tab** of the **Options dialog (Section 3.1.1.4.3)**.

3.2.2.5 Good data end at radius =



This setting determines the upper radius limit on raw scan data to be fitted. This setting can be used to exclude or limit the region near the cell base where solutes are accumulating. The upper radial limit can be altered either by directly entering a number in the text box, or by dragging the right-hand green vertical marker displayed on the 2D graph to a new position.

The extent to which the buildup of solutes needs to be excluded depends in part on which fitting function will be used to fit the data (see **Fitting function options (Section 3.2.3.3)**). Some fitting functions do not account for the cell base (they treat the cell as semi-infinite), and thus the region near the cell base must be excluded. Whether the default fitting model includes buildup at the cell base can be set within the **Fitting Model Defaults tab** of the **Options dialog (Section 3.1.1.4.3)**.

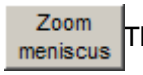
The default fitting function does account for buildup at the cell base, but nonetheless it is usually desirable to exclude at least a portion of the data near the base. For a more thorough discussion of these issues see **Should I fit the region near the cell base? (Section 5.1.3)**, but briefly here are several reasons why some data exclusion may be needed:

1. This region is subject to optical artifacts arising from reflections from the cell base and/or the steep concentration gradients of solutes and/or buffer components which are common there.
2. The concentration near the base may exceed the linear range of the optical system. This problem can also be avoided by using the **Limit maximum buildup at cell base (Section 3.2.2.8)**.
3. the concentration near the base often reaches levels where solution non-ideality effects are very strong, and those effects are not included in the fitting models

4. for most standard centrifuge cells the radial position of the cell base is not uniquely defined because the cell bottom is flat rather than being curved along a specific radius

The default radial limit is controlled by a setting within the **Cell & Data Options tab** of the **Options dialog (Section 3.1.1.4.3)**.

3.2.2.6 Set meniscus & data region - Zoom meniscus button

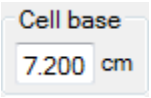


This button will zoom the 2D graph to show only a radial region 2-3 mm wide centered around the current setting of the meniscus position. This is intended to help position the meniscus more accurately.

This command does not alter the Y axis scales and thus it may still be necessary to manually zoom with the mouse to increase the vertical magnification.

This is a toggle button and will remain down to alert you that the graph is currently in zoomed meniscus mode. You can unzoom either by toggling the button off or by hitting the Z key while the graph has the *focus*.

3.2.2.7 Cell base position



The position of the cell base can be set either by entering a value into the text box shown above, or by clicking the blue vertical marker shown on the graph and dragging it to a new position.

The cell base parameter is relevant only when the fitting function models the buildup of solutes at the cell base. That is, it is only relevant for the **Behlke function with base** choice within the **Select fitting model - Fitting function options (Section 3.2.3.3)** on the **Select fitting model page (Section 3.2.3)**. For all other fitting functions the cell base position has no effect on the fit.

Tip 1: If your data are significantly influenced by the buildup at the cell base then it is important that you either:

1. exclude those regions that are influenced by this buildup, and then select a fitting function that does not include buildup at the cell base, or
2. select the **Behlke function with base** fitting function and properly set the cell base position (and you may need to also treat that position as a fitting parameter).

Tip 2: If you alter the default cell base position the program will assume this position has some significance and hence will select the **Behlke function with base** fitting function if that was not already selected. A message box will be displayed if this automatic switch occurs.

The default position of the cell base can be set within the **Cell & Data Options tab** of the **Options dialog (Section 3.1.1.4.3)**. Whether the default fitting model includes buildup at the cell base can be set within the **Fitting Model Defaults tab** of the **Options dialog (Section 3.1.1.4.3)**.

3.2.2.8 Limit maximum buildup at cell base

When the **Limit checkbox** is checked all data points above a specified concentration (in signal units) will be excluded from the fit. The concentration limit can be specified either by typing an entry into the text box or by clicking and dragging the horizontal dotted yellow line on the graph that marks the upper limit.

There are several reasons why it is often desirable to cut-off the fitted data near the cell base to limit the extent of solute buildup:

1. to keep from exceeding the linearity of the absorbance optics
2. to exclude data where solution non-ideality effects are large
3. to exclude data at concentrations where components are irreversibly aggregating or precipitating.

For a more thorough discussion see **Why should I limit the buildup at the cell base? (Section 5.1.2)**.

When this option is used the effect is to cause the upper radius limit of the fitting region to move further to the left as the buildup increases. However for the early scans (before there is any significant buildup) the limit will be determined (as usual) by the **Good data end at radius = (Section 3.2.2.5)** setting.

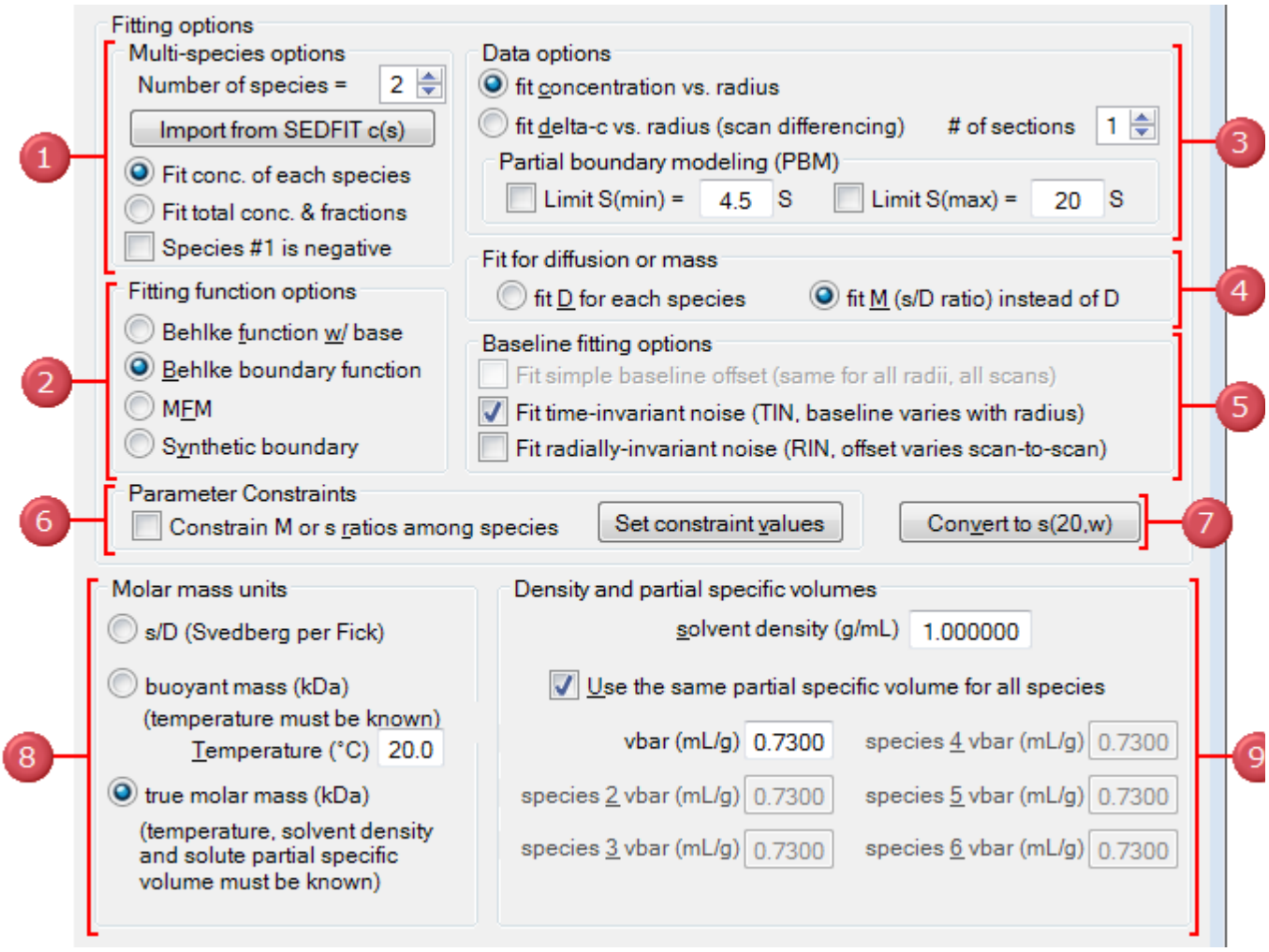
When the **Limit checkbox** is first checked a program wizard makes an estimate of an appropriate limit.

- For all optical systems the wizard first sets a limit that is a multiple of the estimated loading concentration. That default value is set as 200% of the loading concentration when the program is installed, but can be altered on the **Cell & Data Options tab** of the **Options dialog (Section 3.1.1.4.3)**. Note that for data where the baseline is not flat across the cell (data other than true absorbance) the estimate of the loading concentration may not be very accurate.
- For absorbance data a program default sets a maximum OD that will be retained, which may over-ride and reduce the value calculated above based on the loading concentration. That default maximum OD limit is set as 1.4 OD when the program is installed, but can be altered on the **Cell & Data Options tab** of the **Options dialog (Section 3.1.1.4.3)**.

3.2.3 Select fitting model page

This page is used for selecting the theoretical model which will be used to fit the data (number of species, type of fitting function, choices for removal of systematic noise), as well as certain options which affect the way the data and fitting parameters are treated.

Each control or group of controls on the image below is a clickable *hotspot* that will lead to further information, or just use the numbered list below the image.



Below are brief descriptions of the controls on this page; more details can be found by following the links.

1 **Multi-species options (Section 3.2.3.1)** - This section allows you to manually set the number of species (allowable range 1-6) with the up/down spinner, or to set the number of species and the properties of those species using a *c(s)* distribution you have previously calculated for this sample. When more than one species is being fitted it provides options to fit their individual concentrations or instead to fit the total concentration and the species fractions.

2 **Fitting function options (Section 3.2.3.3)** - Choose between fitting functions which include the buildup of solutes at the cell base, or one which does not. The MFM function is provided for backward compatibility with earlier program versions. The synthetic boundary function is used only for synthetic boundary cells.

3 **Data options (Section 3.2.3.4)** - You may fit either to the raw scans or to data from scans which have been subtracted in pairs to remove the time-invariant noise ('scan differencing' or 'delta-*c* mode'). Options are also provided which allow you to exclude species from the fit based on minimum and/or maximum sedimentation coefficient values (partial boundary modeling).

4 **Fit diffusion or mass (Section 3.2.3.6)** - Choose either *D* or molecular mass *(s/D* ratio) as the second hydrodynamic parameter used to characterize each species.

5 **Baseline fitting options (Section 3.2.3.5)** - Options to remove systematic noise from the raw data via fitting a simple baseline offset, a full baseline profile (TIN), and/or offsets which vary from one scan to the

next (RIN or *jitter*).

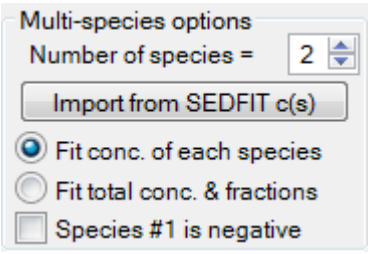
6 **Constrain *M* or *s* ratios among species (Section 3.2.3.7)** - Use constraints between the properties of different species to force them to behave as expected for a series of oligomers or ones with the same hydrodynamic shape (f/f_0 ratio).

7 **Convert to s(20,w) (Section 3.2.3.10)** - Report fitted sedimentation and diffusion coefficients on the standardized scale rather than in 'raw' experimental units.

8 **Molar mass units (Section 3.2.3.8)** - Choose among true mass, buoyant mass, or *s*/*D* ratio. Only relevant when the **Fit M (s/D ratio) instead of D (Section 3.2.3.6)** option is on.

9 **Density and partial specific volumes (Section 3.2.3.9)** - Set the solvent density and sample partial specific volume(s). Only relevant when the **Fit M (s/D ratio) instead of D (Section 3.2.3.6)** option is on).

3.2.3.1 Select fitting model - Multi-species options



This section includes an up/down spinner to set the total number of species, radio buttons to select how the amounts of different species are fitted, the **Import from SEDFIT c(s) button (Section 3.2.3.2)** to initialize a multi-species fit using data from a *c(s)* distribution that was calculated by SEDFIT, and a check box which allows the signals for species #1 to be negative.

Number of species up/down spinner

You may select 1 to 6 species by either typing the number into the text box or using the up and down arrows on the spinner control to increment the number up or down.

It is important to keep in mind that for multi-species fits

1. the different species are assumed to sediment independently (behave as a simple mixture)
2. the different species are assumed not to inter-convert over the time course of the velocity run

Thus this program is generally not appropriate for analysis of interacting systems unless the rates of the association-dissociation process are very slow and the sample is effectively a mixture.

Further, the fact that the program implements six species is not meant to imply that you will be successful at fitting that many to your data! To do so it would likely be necessary to impose **constraints (Section 3.2.3.7)** on their masses or sedimentation coefficients

Fit total concentration and species fractions or fit concentrations for each species

For multi-species fits the fitting parameters for the species concentrations can be organized in two slightly different ways:

1. directly fit the concentrations of each species, with the loading concentration implicitly determined by their sum

2. directly fit the loading concentration, and then fit the species fraction for species 2 through N (the species fraction for species 1 is then implicitly defined as (1 - the total fraction of all other species).

Which option is better depends on which information is most useful to the analyst, and also in part on the nature of the sample. If the fractions are of more interest than the absolute concentrations, then it is best to fit directly for the fractions, especially because then the confidence intervals will directly measure the precision of the desired result. But if for example you were trying to measure the concentration of a free component that is left over after formation of a complex, then the absolute concentration is probably more useful.

One advantage of fitting for fractions is that this will often produce less cross-correlation between the concentration parameters. A decrease in the concentration of one species can often be partially compensated by a corresponding increase in another (keeping the total concentration constant), so this leads to high cross-correlation.

These radio buttons are only active when the number of species is > 1.

Species #1 is negative check box

This option inverts the sign of the theoretical signals for species 1 to allow for the possibility that a solute is producing a negative signal rather than a positive one (normally all signals and concentrations are assumed to be positive). Usually this option is only needed when some component in the reference channel is producing a significant net signal (its concentration exceeds the concentration in the sample channel). The signals for any other species remain positive, and only the species defined as species #1 can be negative.

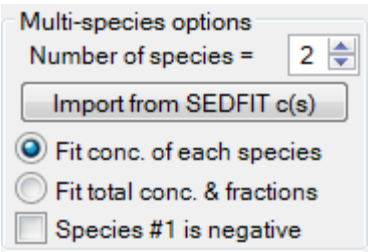
This option cannot be used together with the **Fit total concentration and species fractions** option.

When this option is used the total loading concentration reported in printed reports and on the **Report fit results page (Section 3.2.6)** will be the sum of all species computed with species 1 as a positive concentration.

Tip 1: the fitting of negative signals arising from the reference channel will only work properly if the position of the reference meniscus is quite close to that of the sample meniscus (ideally you should use meniscus-matching centerpieces).

Tip 2: Typically the components creating the negative signals are fairly small molecules, and if so you will want to use the **Behlke function with base** fitting function to model the buildup of material at the cell base correctly.

3.2.3.2 Select fitting model - Import from SEDFIT c(s) button

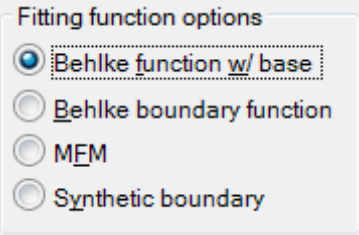


This button allows you to convert the peaks in a *c(s)* distribution into a nearest-equivalent fit as a mixture with the same number of species. This import process does the following things:

1. The concentration and sedimentation coefficient for each species (the initial guesses for the SVEDBERG fit) are determined by integrating the *c(s)* distribution.

2. The mass for each species is estimated from the best-fit *f*/*f*0 ratio, partial specific volume, and density values that were stored in the SEDFIT temporary parameters file.
3. If more than one peak is present, the masses of the different species are linked together (constrained) in a way consistent with a uniform *f*/*f*0 ratio (the same assumption of constant hydrodynamic shape that is used in calculating the *c(s)* distribution). This requires that one species be designated as the 'master', the one whose mass is free to 'float' (vary) during the fit, and then all the others are constrained to have the same *f*/*f*0 ratio as that species (whatever that ratio may change to as the SVEDBERG fit proceeds). A dialog box will open to ask which species should be 'master'; usually the best choice is the one with the highest concentration, and that will be the default selection. See **Constrain M or s ratios among species (Section 3.2.3.7)** for further information about mass constraints.
4. During the integration of the *c(s)* distribution the import wizard checks for half-peaks at the extreme ends of the distribution (the upper and lower sedimentation coefficient limits that were set in SEDFIT). Since such peaks often represent fitting artifacts, the user will be asked whether such peaks should be included in the import.
5. By default the **Fit total concentration & fractions** radio button will be selected.

3.2.3.3 Select fitting model - Fitting function options



This group of 4 radio buttons determines which type of Lamm equation solution is used as the fitting function.

Behlke function with base

This option is in general the best choice, except for synthetic boundary cells. This analytic function is quite accurate for all regions of the cell and for protein masses of ~1 kDa and larger **[ref 5] (Section 7.5)**. When using this function the position of the cell base becomes an important parameter (see

Behlke boundary function

This option **[ref 5] (Section 7.5)** gives a very accurate representation of the moving boundary, but it does not account for the accumulation of solutes at the base of the cell (it treats the cell as semi-infinite in size). The advantage of omitting calculating the buildup at the cell base is that this speeds up the fit roughly 2-fold. Thus this choice is a good one when either:

1. all species in the sample have a fairly high mass and thus the sample does not diffuse very far away from the base, or
2. when the solutes are precipitating at the base of the cell.

Using the **Behlke boundary** fitting function requires that you exclude data from regions of the cell that are being influenced by the cell base by setting the value of **Good data end at radius = (Section 3.2.2.5)** appropriately.

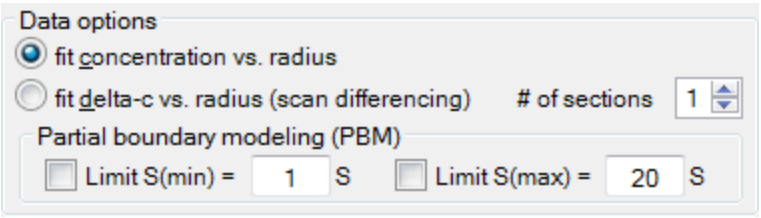
MFM (modified Fujita-MacCosham) function

This option is provided for backward compatibility with earlier versions of SVEDBERG. This function **[ref. 1] (Section 7.1)** is not as accurate as the Behlke functions, and thus not recommended for new analyses. Like the Behlke boundary function choice, this choice does not account for the buildup of solutes at the cell base.

Synthetic boundary

This option provides a function that works for synthetic boundary cells **[ref 3] (Section 7.3)**. These cells are fundamentally different from conventional velocity cells because the solutes can diffuse to radii less than the initial position of the sharp boundary between solution and solvent. Synthetic boundary cells are primarily used for measuring diffusion coefficients, and this function is accurate enough for that purpose, but would probably be not be appropriate for a sample containing multiple components.

3.2.3.4 Data options



This group of controls governs several options how the raw scan data are selected or processed to create the data set to be fitted.

The **partial boundary modeling (PBM)** options are used to remove from the data either fast-sedimenting species (such as aggregates) or slowly-sedimenting species (such as fragments); the **# of sections up/down spinner** is normally only used in conjunction with partial boundary fitting. These advanced options are described in detail in drop-down sections below.

Fit concentration vs. radius

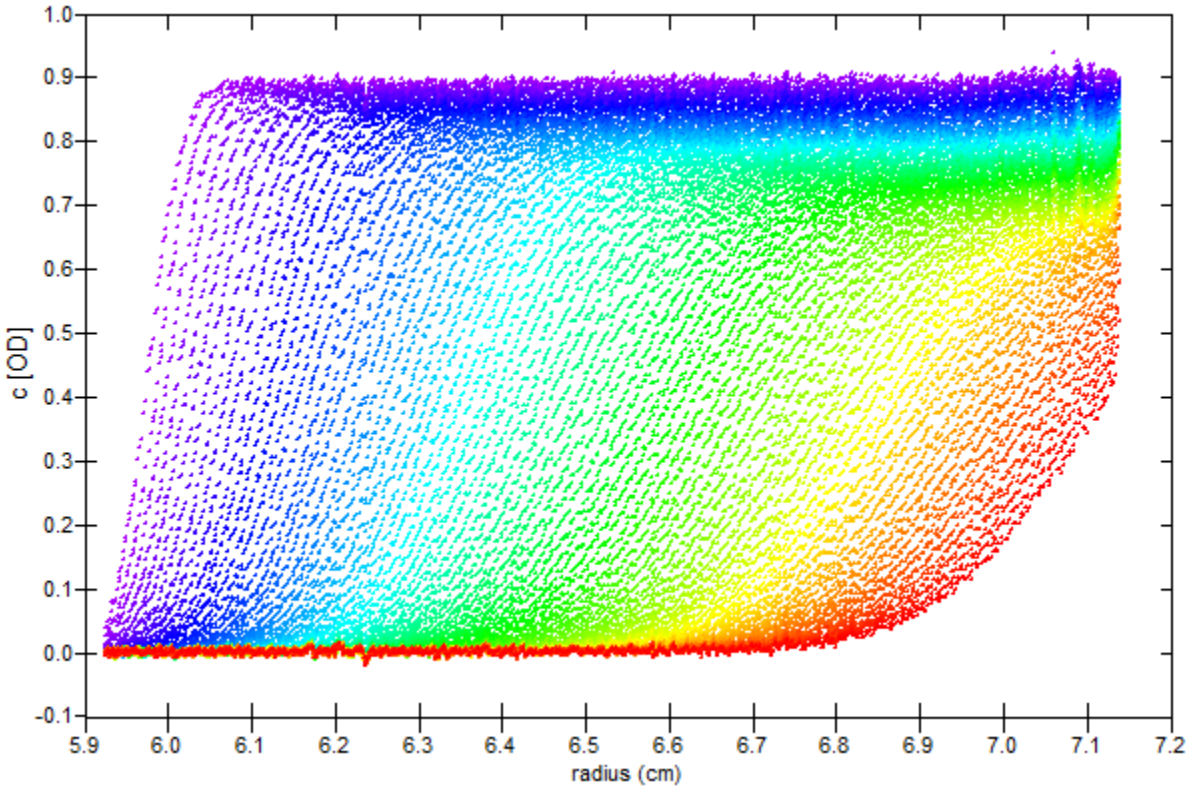
This is the simplest and most conventional choice. The appropriate radial regions of each scan (as determined by your selections on the **Set meniscus & data region page (Section 3.2.2)**) directly become the data points to be fitted, without further selection or processing.

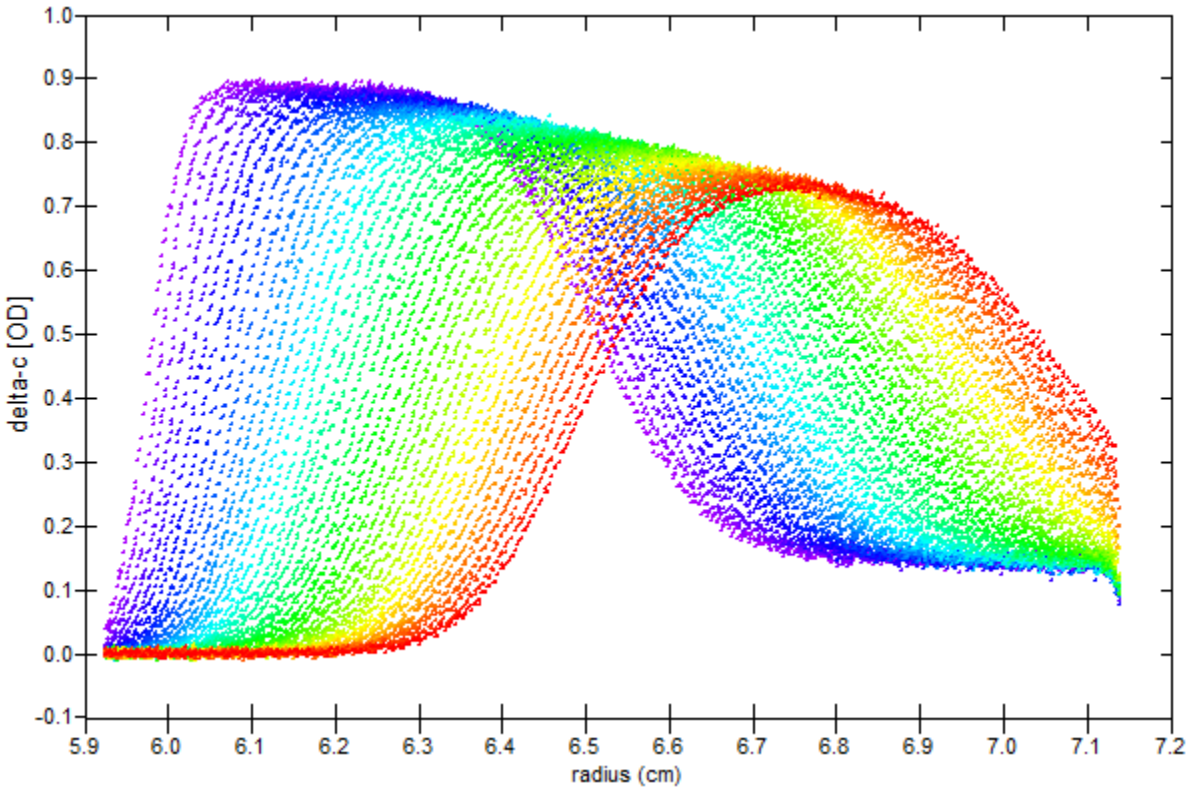
Fit delta-*c* *vs*. radius (scan differencing)

In the delta-*c* (Δc) mode of fitting, rather than directly fitting each individual scan, the scans are subtracted in pairs. This scan differencing removes the time-invariant systematic noise (TIN) such as baseline offsets, window scratches, and most importantly, the window distortions in interference scans via simple arithmetic.

A major advantage of this approach is that the removal of the TIN is model-independent. That is, it does not depend on whether the fitting model actually gives a good description of the data, and does not change when the values of the fitting parameters change.

The following two graphs show the data to be fitted from the same group of absorbance scans arising from either normal or delta-c mode.


c [OD]
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
-0.1
5.9
6.0
6.1
6.2
6.3
6.4
6.5
6.6
6.7
6.8
6.9
7.0
7.1
7.2
radius (cm)


delta-c [OD]
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
-0.1
5.9
6.0
6.1
6.2
6.3
6.4
6.5
6.6
6.7
6.8
6.9
7.0
7.1
7.2
radius (cm)

In delta-*c* mode the data to be fitted typically consist of a series of broad peaks which move to higher radii as the run progresses. Depending on the sharpness of the boundary, and what fraction of the entire run is being fitted, these difference curves can appear similar to the *dc/dt* curves that are used in calculating *g(s*)* distributions.

The default pairing of scans for subtraction is implemented in the same manner normally used in Walter Stafford's "DCDT" method. However, unlike for DCDT, in SVEDBERG there is no restriction that the scans must be closely spaced in time. If there are 10 scans in the fit, then scan #6 is subtracted from scan #1, #7 from #2, etc.

Since this mode automatically removes baseline offsets, upon switching to this mode the baseline offset is automatically set to zero, and fitting of the baseline offset parameter is turned off. It is, however, possible to override this default and still use a baseline offset.

If some of the scans cover different ranges of radii, for each scan pair the range over which the difference is computed is limited to the overlap of the two scans. Thus there is never any extrapolation of the raw data into regions where scanning did not take place. For absorbance scans the radial positions of individual data points differ from scan to scan, so a simple linear interpolation on the scan being subtracted is performed when needed to provide values at intermediate radial positions.

The **"slope correction" parameter (Section 3.2.4.6)** is not available in delta-c mode.

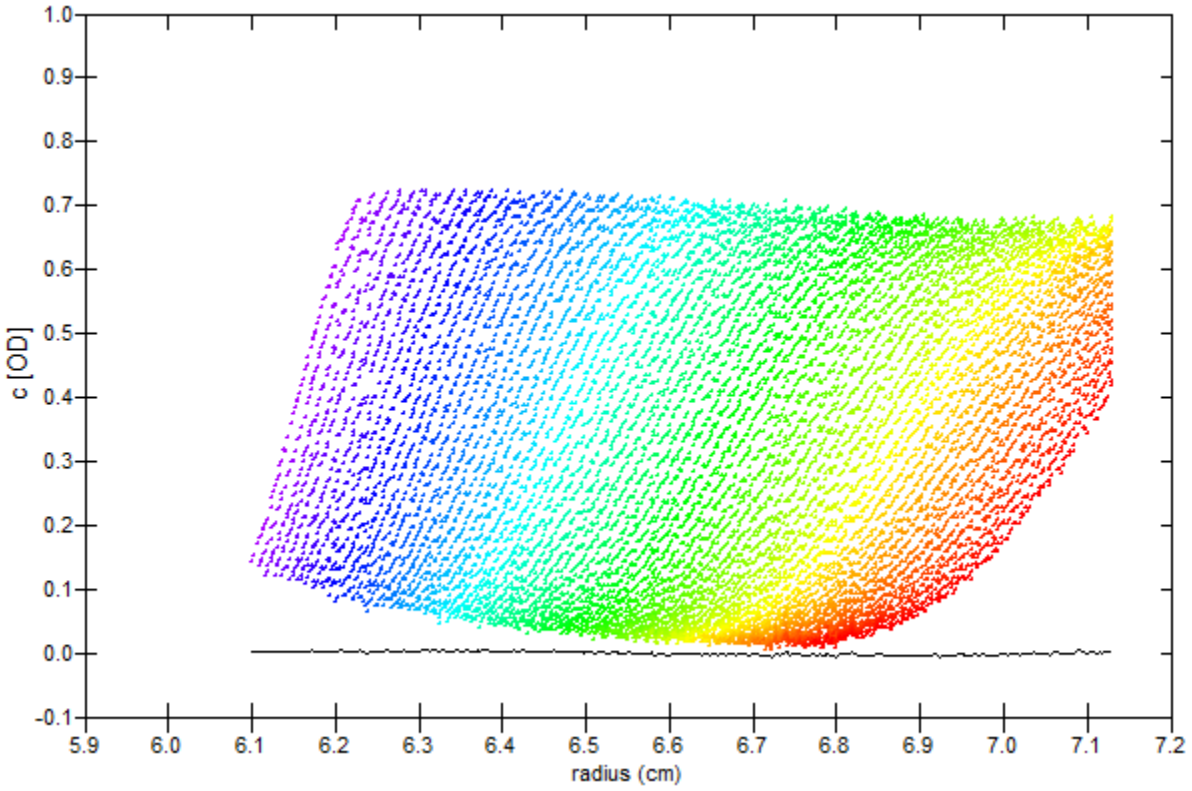
Partial boundary modeling (PBM) options

'Partial boundary modeling' or 'PBM' **[reference 6] (Section 7.6)** is a method for restricting the data being fitted to try to eliminate the influence of species that sediment faster or slower than your range of interest. This is useful, for example, when your sample contains some large irreversible aggregates, or a lower mass impurity, and you want to obtain the information about the major component(s) without trying to explicitly model the amount and properties of the "extra" or "uninteresting" components. That is, you could determine whether your new protein exists in solution as a monomer, or instead as a higher oligomer, by fitting the main part of the boundary, while ignoring the fact that this preparation is not completely pure and homogeneous.

The exclusion on the basis of sedimentation coefficients is implemented by reducing the region of the cell that is being fitted for each scan (the fitted region varies from one scan to the next, moving along with the boundary). If the **Limit S(min) check box** is checked then the fitting region for each scan will begin at a radius that corresponds to the boundary position for a species sedimenting at the desired minimum value, s_{min}, is entered into the text box (which is 1 S in the image shown above). For the earliest scans, prior to the time a species at s_{min} has moved away from the meniscus by an amount set by **Set meniscus & data region - Good data end at radius = (Section 3.2.2.5)**, the starting radius for fitting will be determined normally.

Similarly, if the **Limit S(max) check box** is checked then the fitting region for each scan will end at a radius that corresponds to the boundary position for a species sedimenting at the desired maximum value, s_{max}, is entered into the text box (which is 10 S in the image shown above). For the later scans, for which a species at s_{max} has reached the bottom of the cell, the ending radius for fitting will be determined normally, by the position specified by **Set meniscus & data region - Good data end at radius = (Section 3.2.2.5)** or by **Set meniscus & data region - Limit maximum buildup at cell base (Section 3.2.2.8)**.

Both of these options can be combined to force the fit to narrowly focus on a single boundary (which normally would be the main boundary). The graph below shows the data to be fitted when this approach is applied to the same experiment shown in the graphs above illustrating **delta-*c* mode**. The sedimentation coefficient of this main species is 6.3 S, and the data range has been limited to cover only species between 4.7 and 7.9 S.



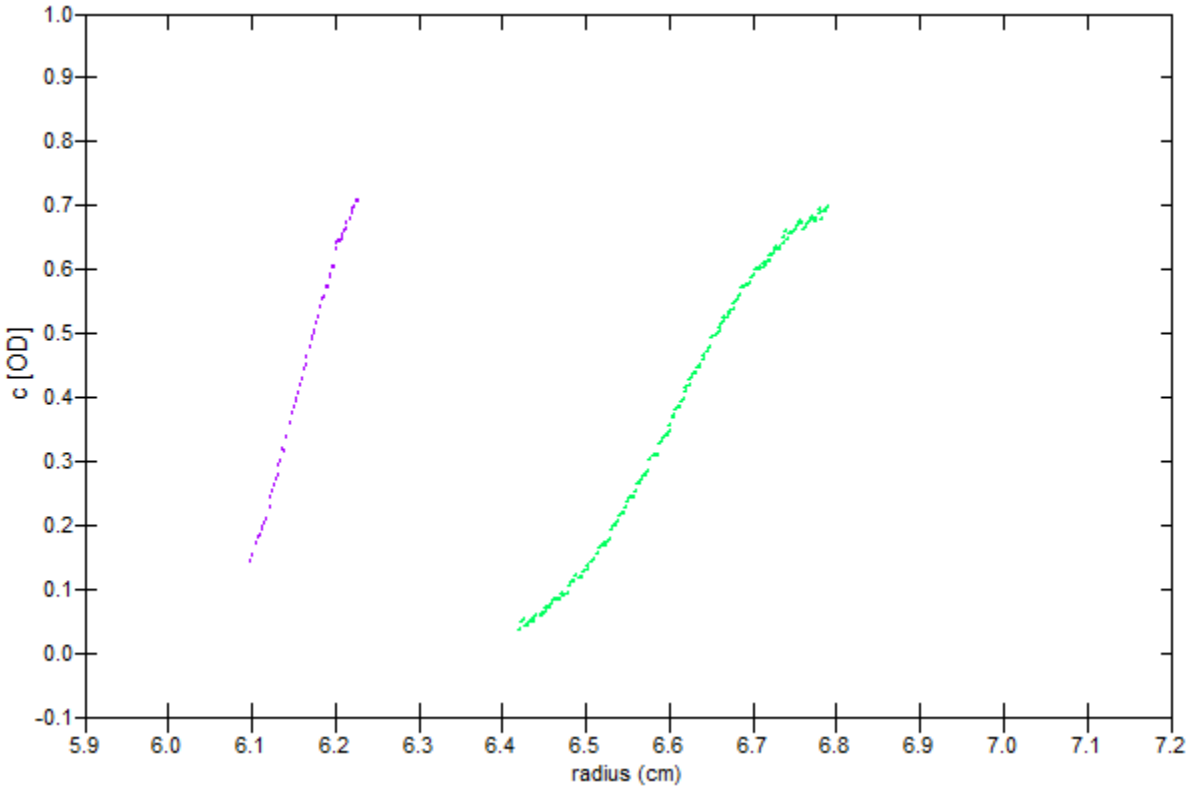
Note that **for PBM fits you usually do not want to load the scans early in the run** (the first ~20% of the total boundary movement), as has been done in this example. One reason is that the resolution between species with different sedimentation coefficient is poor early in the run, so the sedimentation coefficient limits will not be successful at excluding the species you are trying to exclude. A second reason is that trying to fit the TIN can create significant statistical anomalies for PBM fits, and those problems are worse for scans early in the run (because the number of scans contributing data at each radius is small). See **Reference 7 (Section 7.7)** for further details and discussion of this issue.

# of sections option for delta-c mode

Removal of time invariant noise (TIN) can be problematic for fits which use **partial boundary modeling (PBM)**. Directly fitting the TIN for such fits can result in statistical anomalies, and the fact that the computed TIN varies as the parameter values change (the TIN is model-dependent) can result in poor precision for the fitted parameters (very broad confidence intervals), as shown in **Reference 7 (Section 7.7)**. Therefore the model-independent removal of TIN via scan differencing (**delta-*c* mode**) is especially desirable for PBM fits.

For delta-*c* fitting the **# of sections up/down spinner** is normally set to 1, which gives a pattern of pairing the scans to be subtracted as described above. However this pattern means that the scans being subtracted are usually quite far apart in time (the boundary has moved a considerable distance), which creates problems when combining delta-*c* mode with PBM. The only radial regions that can be used for delta-*c* mode subtraction are the regions which are common (overlapping) for the pair being subtracted. However for PBM fits the fitting region moves along with boundary, so if you try to subtract a scan late in the run from one early in the run, then there may be zero (or very little) overlap, especially if the range of sedimentation coefficients being included in the fit is narrow (to isolate a single boundary).

To illustrate, consider the PBM fit shown above in the PBM drop-down section, where only a range from 4.7 to 7.9 S is being fitted. That PBM fit used 70 total scans. If you tried to do that fit in standard delta-*c* mode, this would pair the 1st and 36th scans in that group, which correspond to the data shown in the graph below. Clearly there is no radial overlap between these scans, so in standard delta-*c* mode there would be no data left to fit!

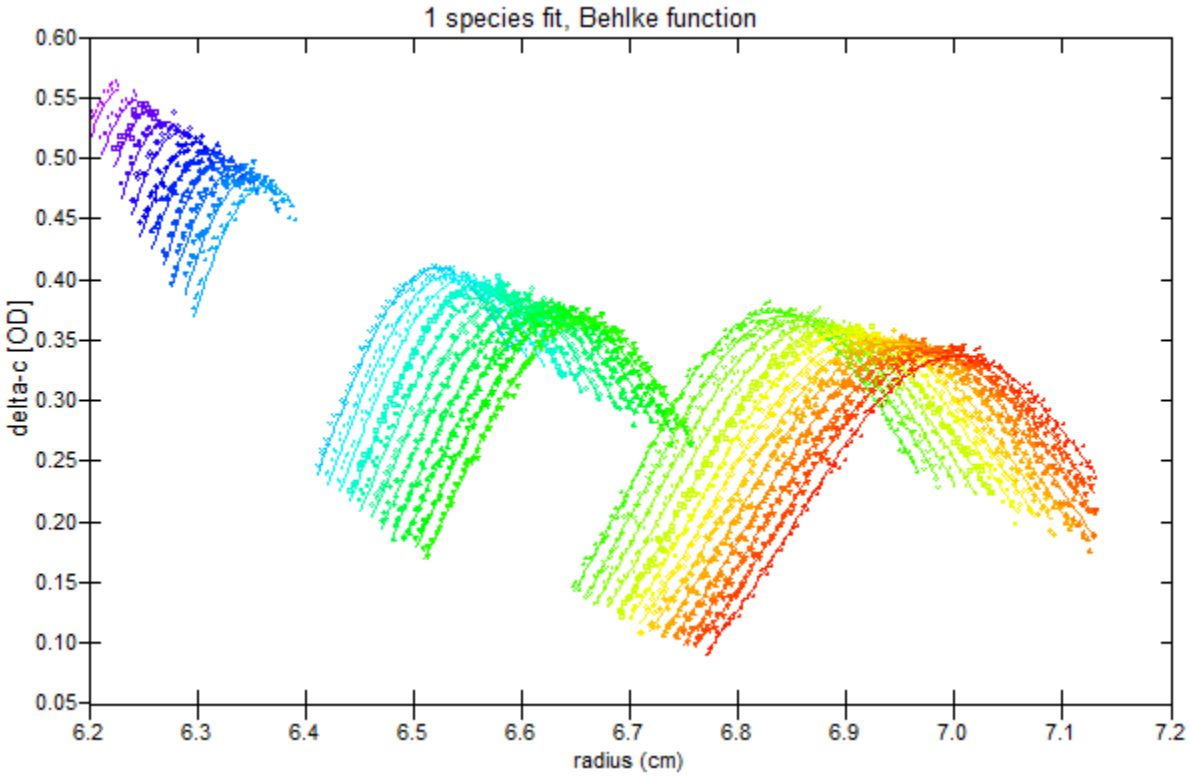


One way around this problem is to alter the scan differencing pattern by dividing the scans into several sections and then do scan differencing within each section. In the standard algorithm when for example a total of 16 scans are being analyzed scan #9 is subtracted from scan #1, scan #10 from scan #2, … and scan #16 from #8. If instead these 16 scans were first divided into two continuous sections, and the usual pattern then applied to each of those two sections, one could pair the scans (5,1) (6,2) (7,3) (8,4) (13,9) (14,10) (15,11) (16,12). This reduces the boundary movement between the scans being paired by approximately 2-fold, thus greatly increasing the region of radial overlap.

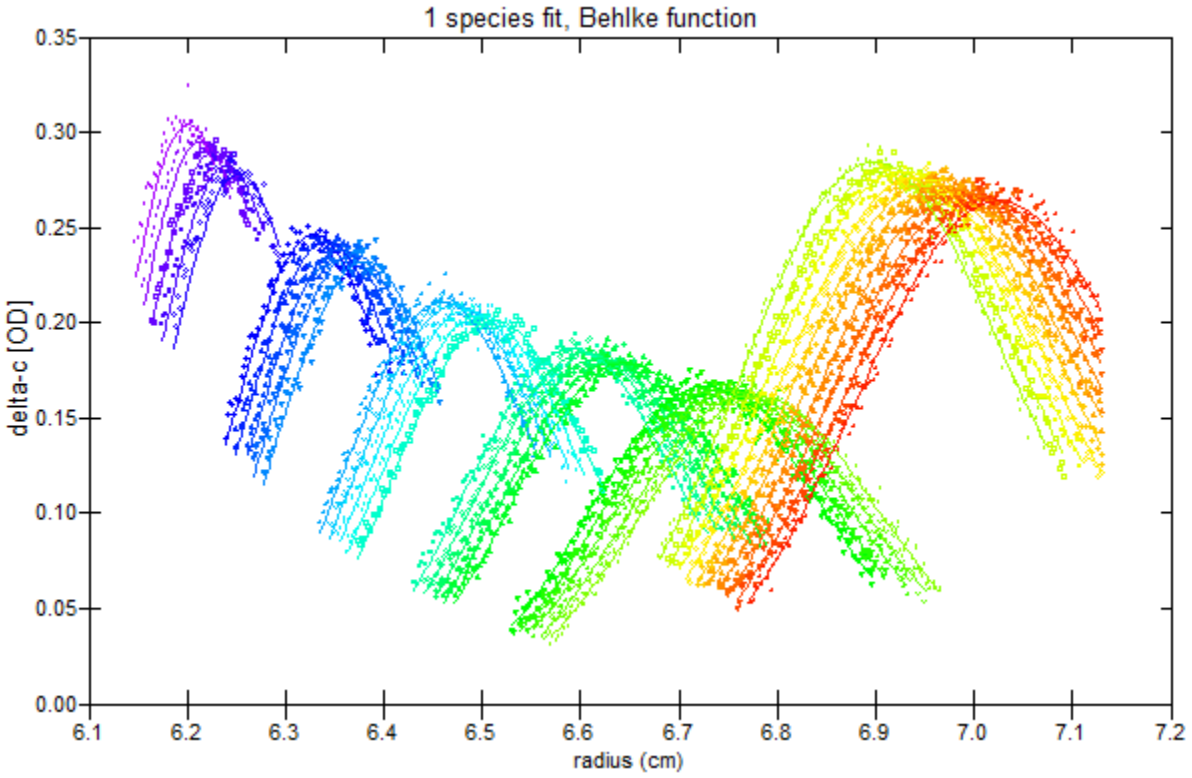
The number of sections can be increased further (and actually at least 3 sections are needed to produce overlap for the example shown above), and in theory can increase until it produces a minimum of 2 scans (1 pair) per section. This approach is not restricted to even numbers of sections, and can be applied to arbitrary total numbers of scans, but in the general case one section will have a larger number of scan pairs than the rest. In such cases, the scans are grouped such that the section with the most scans is the last one (to maximize the overlap of the early scans). The use of multi-section scan differencing was introduced in **Reference 7 (Section 7.7)**.

However the key drawback of using more sections is that the magnitude of the scan differences (Δc) drops substantially (roughly inversely with the number of sections), but the photometric noise remains constant. Thus using multiple sections produces a trade-off: more sections increase radial overlap, which results in larger numbers of data points being included in the fit, but those data points will have a lower signal/noise ratio.

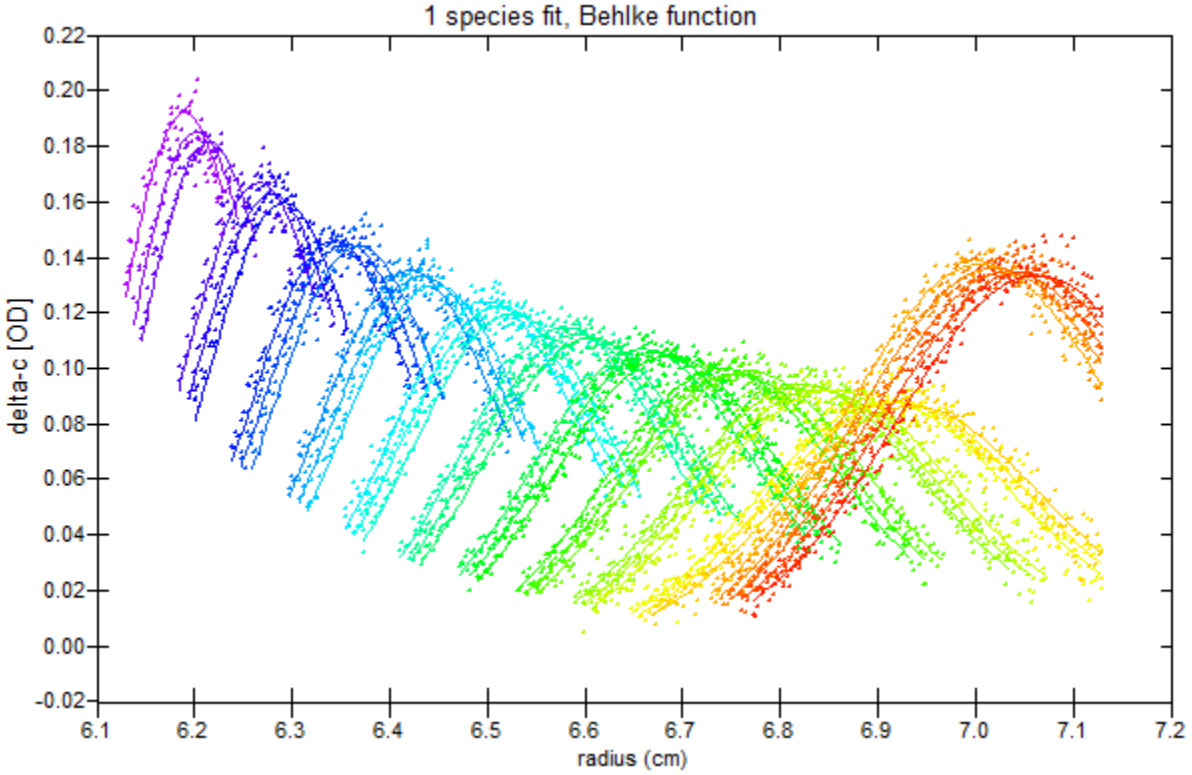
For example, if we use 3 sections with our example experiment, be is an overlay of th data being fitted (points) with the best fit (lines). The 3 sections give enough overlap to produce 3,316 data points to be fitted.



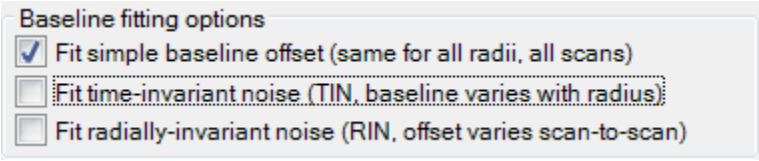
This next graph illustrates using 6 sections with these same data. You can see that the greater overlap gives more data points, especially for the early scan pairs (the violet and blue curves), so there are now 4,295 data points to be fitted, but as you see the amplitude of the delta-c values is about 2-fold lower.



Pushing up to 11 sections gives the results shown below. The noise in each delta-*c* curve is now much more evident, and the number of fitted points only increases to 4,738.



3.2.3.5 Baseline fitting options


Baseline fitting options
Fit simple baseline offset (same for all radii, all scans)
Fit time-invariant noise (TIN, baseline varies with radius)
Fit radially-invariant noise (RIN, offset varies scan-to-scan)

This group of 3 check boxes controls how the baseline for the cell (the measured signal at zero concentration of macromolecule) is handled in the fitting, and also whether the fitting allows for vertical shifts of the data from one scan to another ("jitter", radially-invariant noise or RIN).

Fit simple baseline offset

The **Fit simple baseline offset** option allows for a constant offset (the same value for all radii and all scans). This may be sufficient to handle absorbance data (at least when the windows are clean and free of scratches), and may be needed when the sample or reference contains some small molecule that has significant absorbance but does not sediment. The value of the offset is usually treated as a fitting parameter, but could be fixed if it is known independently (for example from a scan at the end of the run, after all the macromolecules have pelleted).

Note that while the use of a baseline offset is allowed in **delta-c mode (Section 3.2.3.4)**, scan differencing should in principle complete remove offsets and thus usually for **delta-c mode (Section 3.2.3.4)** the offset should be kept fixed at zero.

Fit time-invariant noise (TIN)

The **fit time-invariant noise** option will include in the fit a baseline profile (a zero offset which varies as a function of radius), using the approach developed in **reference 8 (Section 7.8)**. Rayleigh interference scans nearly always show a baseline profile which is not uniform across the cell (a horizontal line), and fluorescence scans may also do so (especially when the fluorophore is sticking to the cell windows). Further, while the baseline for absorbance scans is usually *fairly* flat (independent of radius), it is typically not completely flat even when the cell windows are perfectly clean and free of scratches. Thus including a fitted TIN will nearly always produce significantly lower rms residuals.

The TIN itself is calculated at a set of radial values with uniform spacing. The radial point spacing is set at the average value for the scans and radial regions being fitted (this will correspond to the experimental point spacing for interference and fluorescence scans, where the spacing is uniform). Since the recorded radial positions vary from one absorbance scan to the next, the computed TIN is interpolated to give the value at the recorded position when that is necessary.

Tip: Fitting the TIN is usually not recommended unless you are fitting at least ~50 total scans. To uniquely establish the TIN in each region of the cell also requires multiple replicate scans with no boundary movement in that region. Therefore when fitting the TIN the scans you select for fitting should always include the entire range of boundary movement.

While fitting the TIN has a number of desirable properties, it also has some drawbacks. Including the TIN adds hundreds of additional fitting parameters (up to ~1600 for interference scans). Fortunately these are linear parameters for which the best-fit values can be determined algebraically, but nonetheless they are additional fitting parameters and that can have undesirable statistical consequences, including in some cases substantial reductions in the precision of the hydrodynamic parameters and/or certain raw data points having greatly different weight (influence on the fit) than others (**ref. 7 (Section 7.7)**).

An important point to remember is that the fitted TIN is always model dependent. That is, the computed TIN

will only be correct if the fitting model provides a perfect explanation for the boundary movement, and the resulting residuals are completely random (no sources of systematic noise remain unaccounted for). In contrast, removal of the TIN via scan differencing (**delta-c mode (Section 3.2.3.4)**) is model independent---it is completed fixed once the scans are recorded and selected for analysis, and remains the same no matter whether the fit is good or bad.

For a further discussion of the merits and drawbacks of different approaches see **Which method of systematic noise removal should I use? (Section 5.1.4)**

Caution: Fitting the TIN when also using the **partial boundary modeling (PBM) options (Section 3.2.3.4)** can create undesirable statistical anomalies. See **Reference 7 (Section 7.7)** for details and discussion of this issue.

The actual number of points used in calculating the TIN is reported within the **Fixed, constrained, & implicit parameters (Section 3.2.6.2)** table on the **Report fit results page (Section 3.2.6)** as well as in the printed reports. This value is also included whenever the *degrees of freedom* for the fit are reported (as it should be since these are fitting parameters), but note that this is not done in SEDFIT or SEDPHAT. After the TIN has been calculated the table of values can be copied to the Clipboard or written as a text file using the **Edit menu (Section 3.1.1.2)**, or written out as a scan file from the **File menu (Section 3.1.1.1)**.

Fit radially-invariant noise (RIN)

The **fit radially-invariant noise** option will include in the fit a zero offset (vertical displacement of the raw data) which varies from one scan to the next, but which remains constant across the cell. The RIN is handled as a set of offset parameters (one per scan) using the approach developed in **reference 8 (Section 7.8)**. Rayleigh interference scans nearly always show small (fraction of a fringe) but significant vertical shifts between scans (which goes by the name of 'jitter' as well as 'RIN'). Pseudo-absorbance scans also usually show significant vertical displacements arising from intensity drifts and/or sudden shifts during the course of the run.

TIP: Including a fitted RIN is not recommended for absorbance or fluorescence scans. A properly functioning instrument should not have absorbance or fluorescence offsets from one scan to the next. Further, a major advantage of these optical systems over interference scans is that zero absorbance or fluorescence intensity has a physical meaning, whereas the zero for interference scans is always arbitrary. Using a fitted RIN with absorbance or fluorescence data gives up this advantage.

The removal of this type of systematic noise can be handled either (1) via fitting the RIN or (2) arithmetically by forcing all the scans to align vertically at some position in the cell (over a small range of radii) using the **Jitter removal options (Section 3.2.2.2.1)**. These two approaches are in principle mutually exclusive. However since by default jitter is always removed by the **jitter wizard (Section 6.5)**, the program allows RIN fitting to be used in addition, but when both are used together then in reality the scans may no longer align vertically over the matching region. Note also that the distinction between *fringe jumps* and 'jitter' becomes moot when the RIN is fitted (the RIN can include both the integer and fractional portions).

An important point to remember is that the fitted RIN is always model dependent. That is, the computed RIN will only be correct if the fitting model provides a perfect explanation for the boundary movement, and the resulting residuals are completely random (no sources of systematic noise remain unaccounted for). In contrast, removal of the RIN via designating an alignment region is model independent---it is completed fixed once the scans are recorded and selected for analysis and the alignment region has been designated, and remains the same no matter whether the fit is good or bad.

For a further discussion of the merits and drawbacks of different approaches see **Which method of systematic noise removal should I use? (Section 5.1.4)**

The actual number of points in the RIN (the number of scans being fitted - 1) is included whenever the *degrees of freedom* for the fit are reported (as it should be since these are fitting parameters), but note that this is not done in SEDFIT or SEDPHAT. After the RIN has been calculated the table of values can be copied to the Clipboard or written as a text file using the **Edit menu (Section 3.1.1.2)**.

Fi

3.2.3.6 Select fitting model - Fit diffusion or mass

Internally SVEDBERG always uses sedimentation coefficient, *s*, and the diffusion coefficient, *D*, as the fitting parameters which characterize the hydrodynamic properties of each species.

However, instead of using *D* as a fitting parameter, it is often desirable to instead use the ratio *s*/*D*, a quantity which is proportional to molecular mass, *M*. In fact, fitting to mass rather than *D* is the initial program default when installed, but that can be altered via the **Options dialog (Section 3.1.1.4.3)**.

One significant advantage of fitting *M* rather than *D* is that this allows using constraints on the mass values for different species (for example fitting to a mixture of monomer and dimer, with the monomer mass as a fitting parameter). See **Constrain M or s ratios among species (Section 3.2.3.7)**.

Why this works

The **Svedberg equation** shows that the *s*/*D* ratio is directly related to molecular mass, *M*:

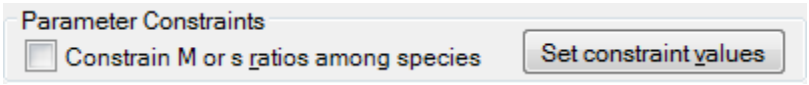
$$M = \frac{sRT}{D(1-\bar{v}\rho)}$$

where *R* is the gas constant, $\bar{v}$ is the partial specific volume, ρ is the solvent density, and T is temperature (Kelvin).

How this works

As stated above, internally the program always uses the value of *D* to determine the shape of the corresponding peak. When you choose to use *M* as the fitting parameter, during each iteration of the fitter the correct *D* value is calculated from the current values of *s* and *M*. Since variations in the *s* value will also alter the apparent mass (the *s*/*D* ratio), it is also necessary to adjust the derivatives of the fitting function to account for that effect.

3.2.3.7 Select fitting model - Constrain M or s ratios among species



Parameter constraints are used with multi-species fits when there is reason to believe two or more species represent oligomers of a common monomer, or have a common hydrodynamic shape (f/f_0 ratio).

The **Set constraint values button** brings up a dialog box that allows the sedimentation coefficients and/or masses of the different species to be linked together and constrained to the ratios expected for oligomers. By constraining these ratios, the number of fitted hydrodynamic parameters is greatly reduced, often resulting in more accurate values for the other parameters that are fitted.

Whenever constraints are currently active, the **Constrain *M* or *s* ratios among species** check box will be checked. Clearing this check box will inactivate all constraints. Setting this check box when no constraints are active will bring up the dialog to set constraints.

The actual dialog box and procedure for establishing the constraints is described in detail below, but first it may be useful to explain the intended use for these constraints.

Why are constraints useful?

As an example, a common degradation pathway for proteins upon storage is the formation of disulfide-linked dimers or higher oligomers, and thus many protein samples could contain a small amount of dimer. If the fraction of dimer is quite small, it will be very difficult for the fit to accurately evaluate its sedimentation and diffusion coefficients (especially *D*).

Therefore if you want more accuracy for the amount and sedimentation coefficient of the dimer fraction (and there is a good reason for believing this is a dimer, such as evidence from non-reducing gels or SEC-MALS) it may be a good idea to constrain the *D* value of the dimer to be consistent with a mass twice that of monomer. That is, if s_1 and D_1 are the current values of *s* and *D* for the monomer, and s_2 and D_2 are the current *s* and *D* values for the dimer, then at every iteration of the fit the program can force the s_2/D_2 ratio to be exactly twice the s_1/D_1 ratio; i.e., $D_2 = s_2 \times D_1 / (2 \times s_1)$.

Similarly, if the fraction of dimer is very small (a few percent) or the signal/noise is poor, it may also be appropriate to constrain the dimer sedimentation coefficients to be an appropriate multiple of the monomer s_1 value. If each monomer is a hard sphere, then theory says that $s_2 = 1.45 \times s_1$ (**Reference 10 (Section 7.10)**). At the other extreme, if two spherical monomers merged to form a spherical dimer, then $s_2 = 2^{2/3} * s_1 = 1.59 * s_1$. Thus the difficulty with constraining sedimentation coefficients is that the ratios depend on conformation.

These constraints can be very useful for hypothesis testing. If we see that a sample fits poorly as a single species, and/or the fit implies a molecular mass significantly less than a known value (broader boundaries), then you may wish to test the hypothesis that the poor result is due to the presence of a small amount of oligomer. This can be done by first constraining both the *M* and *s* value of the second species and fitting. If this gives a good fit and reasonable values for the monomer parameters, then you could conclude that the hypothesis is satisfied. At that point you would probably also want to try releasing the constraint on the oligomer *s* value, re-fitting, and noting whether the fitted dimer *s* value is still consistent with the assumed oligomer size. If this is also true you would have still more confidence that your hypothesis is true. See **Tutorial: Multi-species fits (Section 2.2)** for an example.

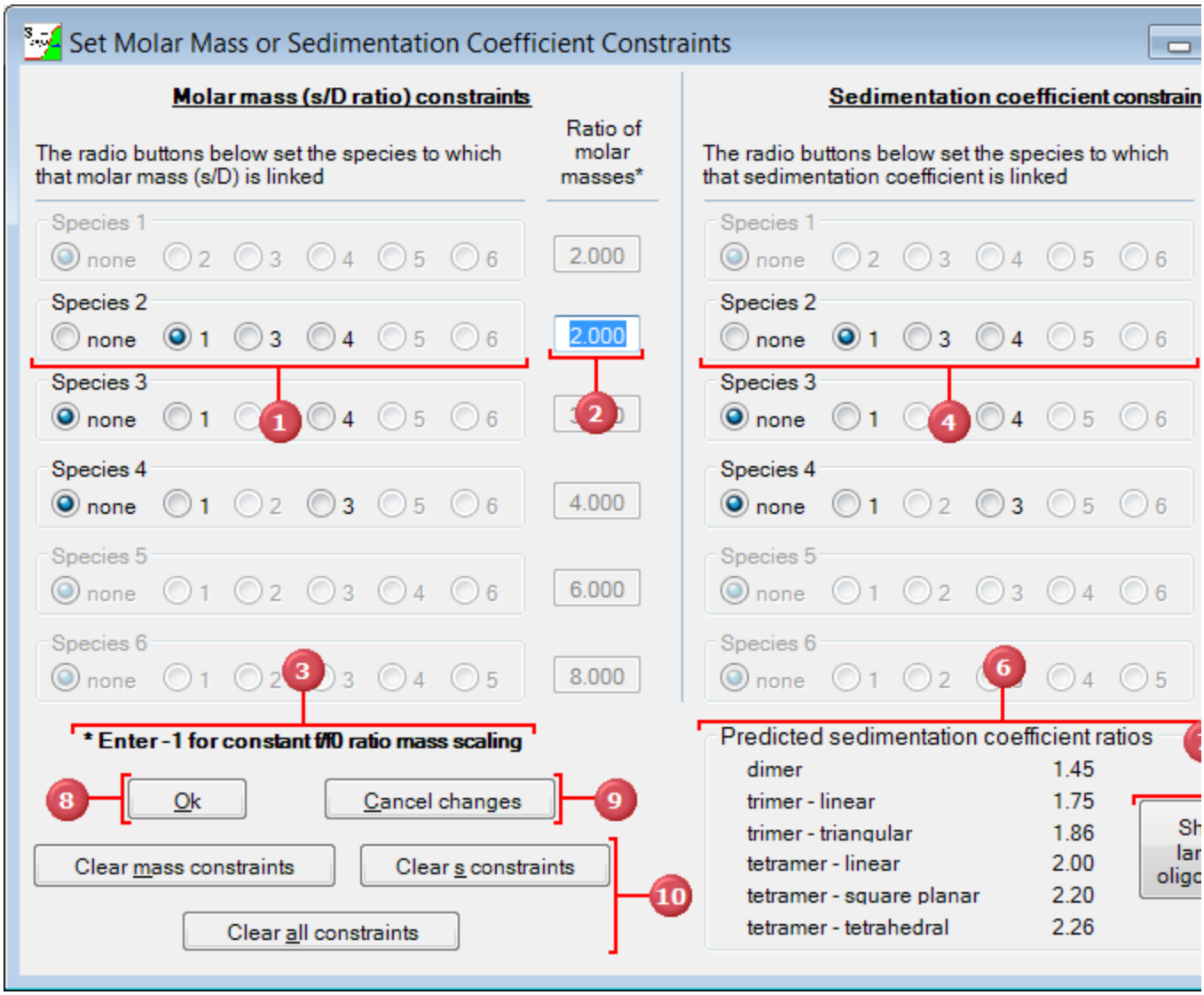
Why are constraints sometimes needed?

The need for constraints arises when:

1. The separation between two species is too poor to uniquely define all the properties of both of them.
2. Only a small amount of a species is present, and either the signal/noise or the separation is too poor to permit uniquely determining all three of its properties (*c*, *s*, and *M* or *D*).
3. You are more concerned about obtaining accurate values for certain properties, but others are less important. The use of constraints (if they are appropriate ones) can often greatly increase the precision of the remaining fitted parameters.

The constraint dialog box:

The dialog box used for establishing or removing the constraints is shown below. In this example, the fitting model includes three species, so the controls related to species 4, 5 and 6 are inactive (grayed).



Mass constraints

1 The image shows an example of a 3-species fit where species 2 is believed to be a dimer of species 1. Hence on the left side the molecular mass of species 2 is set up to be constrained to be exactly twice that of species 1 (whatever that may be).

The desired ratio of masses for the constrained species is entered into the text box 2

Once the constraint link is established to species 1, species 1 cannot itself be linked to other species, so its linking box becomes inactive. As shown, the molecular mass of species 3 is not constrained, but it could still be linked to that of species 1 (perhaps as a trimer or tetramer) and thus the group of radio buttons for linking the mass of species 3 to another species is still active.

3 An alternative type of mass constraint is also available. For **constant f/f_0 ratio mass scaling** the mass of the constrained species is calculated assuming this species has the same hydrodynamic shape (same frictional coefficient ratio relative to an anhydrous sphere, f/f_0) as the species to which it is linked. This constraint is therefore equivalent to the assumption used when calculating the *c(s)* distribution in Peter Schuck's program SEDFIT. When the f/f_0 ratio is constant the mass scales with the sedimentation coefficient to the 3/2 power.

Thus if species 2 is linked to species 1 with this type of constraint we have

$$M_2 = M_1 \times \left(\frac{s_2}{s_1}\right)^{3/2}$$

To use this constant f/f_0 ratio mass scaling simply enter -1 as the mass ratio in the text box (as the reminder on the form says).

Sedimentation coefficient constraints

4 On the right side, the sedimentation coefficient of species 2 is also constrained to be a multiple of that for species 1.

The ratio of sedimentation coefficients for the constrained species is entered into the text box 5.

6 The table at lower right shows theoretical ratios for oligomers of hard spheres. The table has room only for oligomers up to tetramer, but ratios for pentamers, hexamers, and cubic octamer replace that table when the **Show bigger oligomers** button is pressed. 7

In this example species 2 is being constrained at the 1.45 ratio appropriate for a dimer of hard spheres. These ratios for oligomers come from Garcia de la Torre, J. and V. A. Bloomfield (1981), **reference 10 (Section 7.10)**.

Tip: In some cases it may be better to constrain the smaller species to be a sub-multiple of a larger one. For example, if you had a sample where the native protein is a dimer, but it contains a small fraction of a denatured or "incompetent" monomer, it would be best to constrain the monomer mass to be half that of the dimer, and perhaps set the monomer *s* value to be 0.69 that of the dimer (i.e., the inverse of the 1.45 ratio shown in the table).

Other buttons

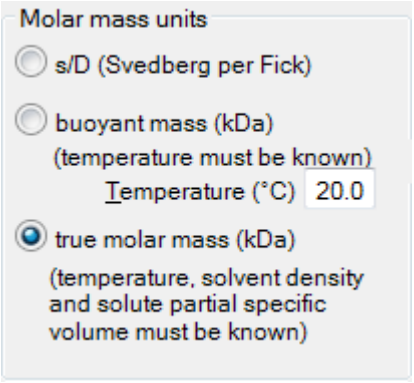
8 When you are done setting constraints the **Ok button** will accept the current settings and return you to the **Select fitting model page (Section 3.2.3)**.

9 If you want to undo any modifications and return to the settings present when this dialog box was opened, use the **Cancel button**, which will close the dialog and return you to the **Select fitting model page (Section 3.2.3)**.

10 This group of 3 buttons provides quick mechanisms to clear out unwanted constraints. The usage is obvious from the names of the buttons.

Note that if you reduce the number of species from the **Select fitting model page (Section 3.2.3)** after constraints are active, any constraints to or from species that no longer exist in the model will be automatically cleared. You will not need to bring up this dialog box to clear them.

3.2.3.8 Select fitting model - Molar mass units


Molar mass units
s/D (Svedberg per Fick)
buoyant mass (kDa)
(temperature must be known)
Temperature (°C) 20.0
true molar mass (kDa)
(temperature, solvent density and solute partial specific volume must be known)

This group of controls allows you to choose among three systems of units for molar mass. This choice is only relevant if you have chosen to use molar mass (the *s*/*D* ratio) as the second hydrodynamic parameter for the fit within the **Fit diffusion or mass (Section 3.2.3.6)** group box.

Generally you will want to use units of true molar mass, and this is the initial program default.

However, in some cases you may not know the values for both ρ and $\overline{\nu}$ and thus cannot calculate the true mass. In other cases you may wish to work directly with the buoyant molar mass or in units of the *s*/*D* ratio (Svedbergs per *Fick*).

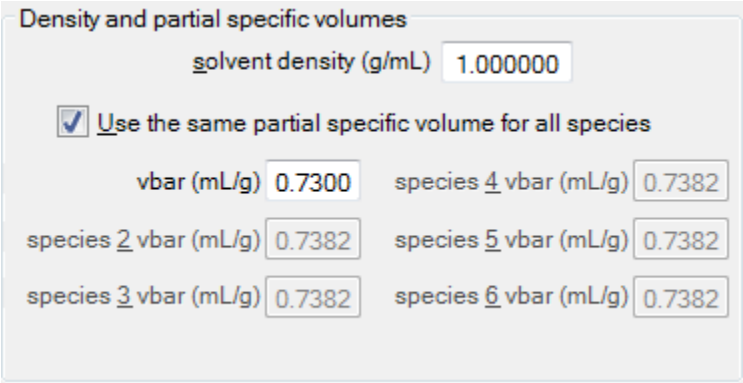
To set the molar mass units:

1. Use the radio buttons to select among the three choices of units.
2. To use units of buoyant mass or true mass, the sample temperature must be known.
   - The program automatically fills in the sample temperature based on the temperature recorded in the header of the last scan used in calculating the *dc*/*dt* data. However, because the temperature may drift during the run you may wish to adjust that value manually. The program internally converts the temperatures from Celsius to Kelvin scales.
   - Please note that changing the temperature on this page overrides the temperature recorded with the scans, and also alters the value used elsewhere in the program (*e.g.* for conversions from raw *s* to *s (20,w)*.
3. If you select true molar mass, go to the **Density and partial specific volumes group (Section 3.2.3.9)** on the lower right side of the page and fill in the values for solvent density and the partial specific volume(s).

To obtain accurate true molar mass you must know the solvent density and solute partial specific volume accurately.

Note that changing the molar mass units on this page will automatically alter the results displayed on the **Report fit results page (Section 3.2.6)** if a fit has been completed.

3.2.3.9 Select fitting model - Density and partial specific volumes


Density and partial specific volumes
solvent density (g/mL) 1.000000
Use the same partial specific volume for all species
vbar (mL/g) 0.7300
species 4 vbar (mL/g) 0.7382
species 2 vbar (mL/g) 0.7382
species 5 vbar (mL/g) 0.7382
species 3 vbar (mL/g) 0.7382
species 6 vbar (mL/g) 0.7382

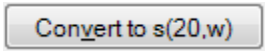
This group of controls is used to set values for the solvent density, ρ, and sample partial specific volume, $\bar{\nu}$. Those values are only relevant and needed when you have chosen to use molecular mass (the *s*/*D* ratio) as the second hydrodynamic parameter for the fit within the **Fit diffusion or mass (Section 3.2.3.6)** group box, and when you have chosen units of true molecular mass within the **Molecular mass units (Section 3.2.3.8)** group box.

- When checked the **Use the same partial specific volume for all species** check box causes the program to assign the same $\bar{\nu}$ to all species (as would be true, for example, if they are all oligomers of a single type of monomer). This is the default state of the program.
  - If your sample contains more than one chemical species, uncheck the box and then fill in the correct values for the other species.
  - If you later turn on the **Use the same partial specific volume for all species check box**, the value entered for species #1 will automatically assigned to all the other species.
- When the program starts, ρ and $\bar{\nu}$ are set to default values. These default values may be altered using the **Options dialog (Section 3.1.1.4.3)**.
- When you alter ρ or $\bar{\nu}$ on this page, those values will also be used in other calculations applied to this **analysis window (Section 3.1.3)**, such as converting raw sedimentation coefficients to *s(20,w)* values.

3.2.3.10 Convert to s(20,w)

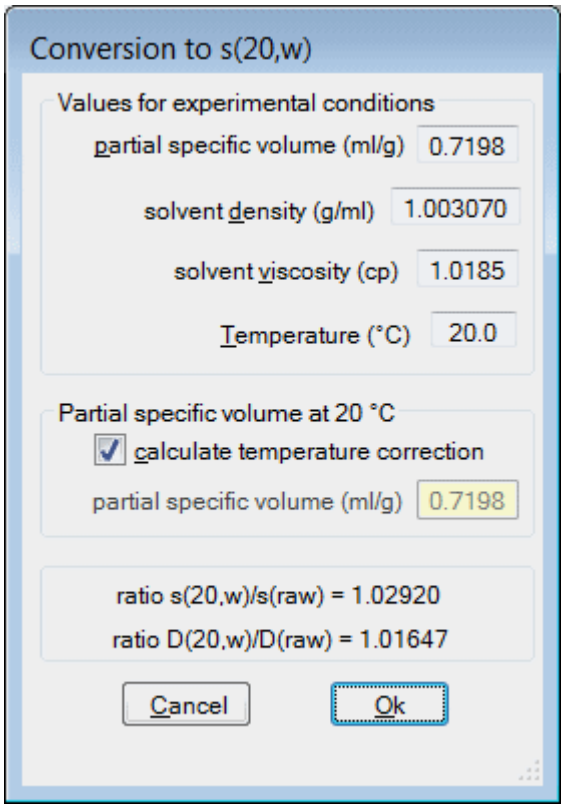
This option will convert the fitted sedimentation coefficient(s) to the $s_{20,w}$ scale.

Note that $s_{20,w}$ will also sometimes be written in this document as *s(20,w)* to avoid problems in generating subscripts



When the program is in "raw" sedimentation coefficient unit mode pressing this button will bring up a dialog box (shown below) for entry of the data needed to perform the conversion (discussed in more detail below).

If this conversion has already been done (as indicated by the button remaining in the down state) then pressing the button will turn off the conversion and revert to raw *s* values.



It is generally desirable to present and report $s_{20,w}$ values rather than raw (uncorrected) sedimentation coefficients. Conversion from raw *s* to $s_{20,w}$ compensates for the factors such as buffer viscosity which are relevant only to a particular experiment, giving the sedimentation coefficient that molecule would have if measured in water at 20 °C. Because of the solvent-independent nature of $s_{20,w}$, it describes quantitatively the fundamental hydrodynamic properties of the molecule and it is this value which is most useful in comparing the sedimentation behavior of different molecules.

The actual scaling factors that will be applied to correct the raw sedimentation and diffusion coefficients to standard conditions are displayed in the panel nearest the bottom of the dialog box.

What about extrapolation to zero concentration?

This program does not attempt to correct for concentration dependence of sedimentation coefficients (extrapolation to infinite dilution) to give $s^0_{20,w}$. Often *dc*/*dt* analysis is applied to samples at concentrations sufficiently low that concentration dependence can be neglected, and if so the corrected results can be considered to be $s^0_{20,w}$ values, but the program will not explicitly designate them as such.

How this conversion affects reported diffusion coefficients

Activating this option will also cause raw diffusion coefficients to be converted to $D_{20,w}$ values, via the formula

$$D_{20,w} = D_{raw} \frac{\eta_{T,b}}{\eta_{20,w}} \frac{293.15}{T}$$

where *T* is the sample temperature (Kelvin).

Data required for conversion

Conversion to $s_{20,w}$ requires knowledge of the solvent density at the experiment temperature, the solvent viscosity, h, at the experiment temperature, the solute partial specific volume (vbar) at the experiment temperature, and the solute partial specific volume at 20 °C, through the relation:

$$s_{20,w} = s_{raw} \frac{\left(1-\bar{\nu}_{20,w}\rho_{20,w}\right)}{\left(1-\bar{\nu}_{T,b}\rho_{T,b}\right)} \frac{\eta_{T,b}}{\eta_{20,w}}$$

where the subscripts refer to the temperature and solvent conditions (b=buffer, w=water). The density and viscosity of water at 20 °C are 0.998234 g/ml and 1.002 centipoise, respectively.

If all those data are not available, or cannot be estimated (*e.g.* by SEDNTERP), then the conversion cannot be made.

The program will optionally automatically compute a temperature correction to the partial specific volume, using the same formula used in SEDNTERP [**reference 4**]:

$$\bar{\nu}_T = \bar{\nu}_{20} + 4.25\times10^{-4}\left(T-20\right)$$

This option is controlled by the check box on the lower frame, and when it is checked the input box for the vbar at 20 °C is disabled (but the computed value is displayed).

The viscosity values are to be entered in centipoise. Since SEDNTERP displays these values in poise, to facilitate copying and pasting numbers from SEDNTERP, whenever a number is entered that is less than 0.1 the program will assume it is a value in poise and multiply it by 100.

The program automatically fills in the experiment temperature based on the temperature recorded in the header of the last scan used in calculating the *dc/dt* data, but since the temperature may drift during the run you may wish to adjust that value manually.

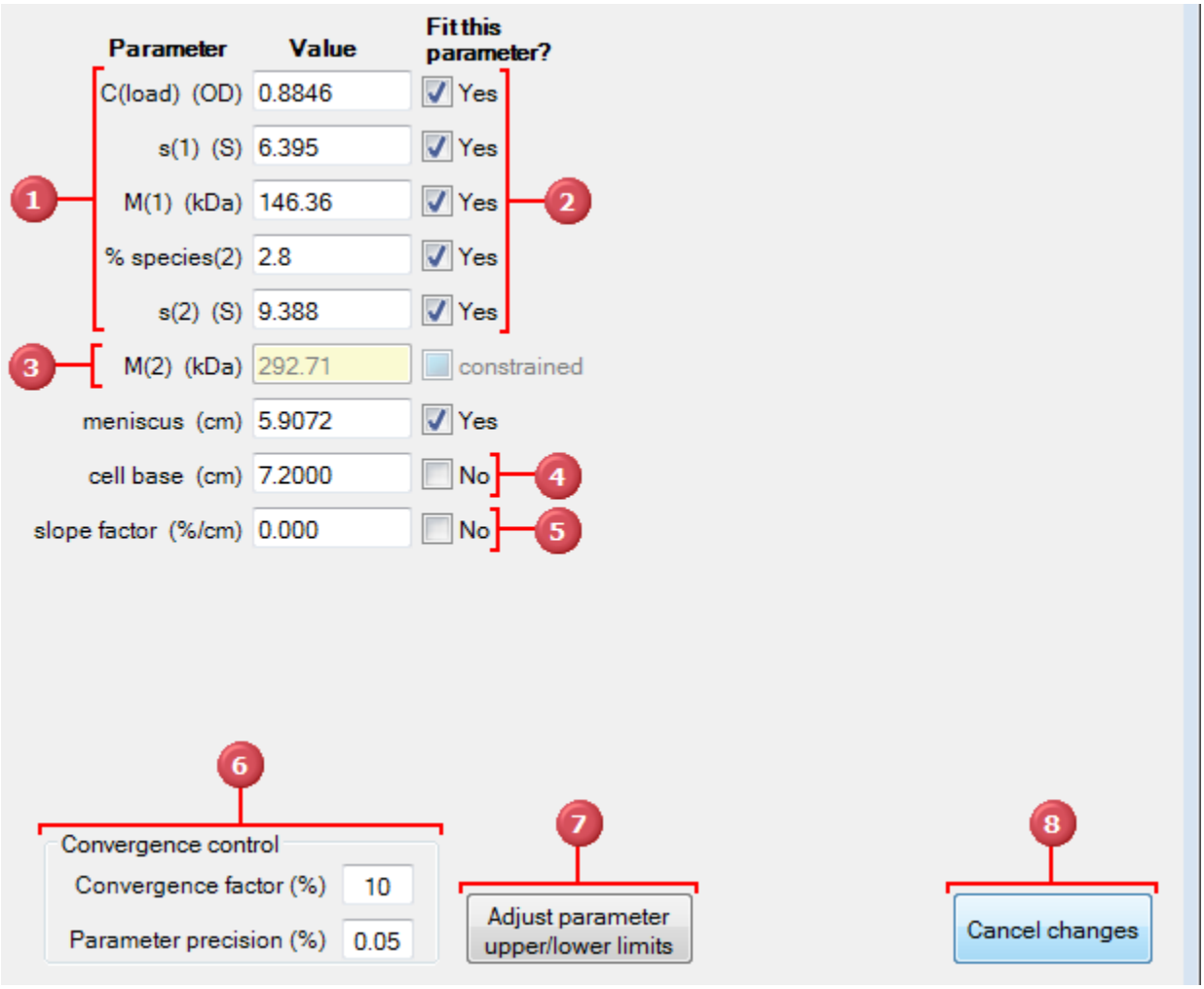
Please note that changing the temperature on this form overrides the temperature recorded with the scans, and also alters the value used elsewhere in the program (*e.g.* for conversions to buoyant or true mass units).

3.2.4 Set/alter fitting parameters page

This page is used to manually enter or alter the initial guesses for the fitting parameters, for specifying whether those parameters will be fitted or held constant, and for specifying the criterion for convergence of the fit.

Each control or group of controls on the image below is a clickable *hotspot* that will lead to further information, or just use the numbered list below the image.

This particular example is that for a fit involving 2 species, and where the molecular mass of species 2 is being constrained to be a multiple of that of species 1.



Below are brief descriptions of the controls on this page; more details can be found by following the links.

1 **Parameter value text boxes (Section 3.2.4.1)** - Input or alter the initial guess (for parameters being fitted) or the desired value for that parameter (for parameters being held constant).

2 **Fit control check boxes (Section 3.2.4.2)** - When checked that parameter will be fitted ("floated"); when unchecked that parameter is held constant ("fixed").

3 **Constrained parameters (Section 3.2.4.3)** - The values for parameters that are constrained are neither fitted nor fixed. Their values cannot be altered from this dialog; they are computed based on the values of other parameters. In the case shown species 2 is constrained to have twice the molecular mass of species 1 (species 2 is a dimer). See **Constrain *M* or *s* ratios among species (Section 3.2.3.7)**.

4 This is an example of a parameter (the position of the base of the cell) that is being held fixed (at 7.20 cm in this case). This is the program default for this parameter. See **Should I fit the cell base position? (Section 5.1.3)**

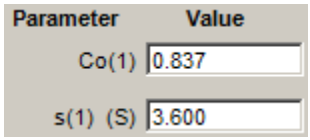
5 The slope factor parameter is an empirical correction for the 'sloping plateaus' artifact of the absorbance optics (see **What is a 'sloping plateau'? (Section 5.1.5)**). When the slope factor is fixed at zero, as in this example, this correction is off; that is the default. This factor should never be fitted or non-zero for anything other than absorbance data.

Convergence control (Section 3.2.4.4) - These two values are used in determining when a fit has converged. The default values work well in most situations. Lower values will produce more iterations, causing the fitter to come ever closer to the best possible fit, but those extra iterations will probably not produce changes that are significant compared to the statistical precision of the parameter values. Higher settings on these values may help the fitter reach convergence on particularly long and difficult fits.

Adjust parameter upper/lower limits (Section 3.2.4.5) - This brings up a dialog box for altering the default upper and lower bounds on each parameter. Generally changes in these values are not needed.

Cancel changes (Section 3.2.4.7) - This is an 'Undo' button that restores all values and settings as they were when this analysis page was first entered.

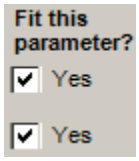
3.2.4.1 Parameter value text boxes



Use these text boxes to input starting guesses for the parameters (or fixed values if that parameter is not being fitted).

- The units for concentrations are OD, fringes, or fluorescence intensity; when fitting **fractions rather than concentrations (Section 3.2.3.1)** the units are percent.
- Sedimentation coefficients are in Svedbergs, diffusion coefficients are in *Fick*s, and masses are either in kDa or Svedberg/Fick (depending on your choice of **molecular mass units (Section 3.2.3.8)**).
- The **Slope correction fitting parameter (Section 3.2.4.6)** is entered as a percent.
- You may enter values in scientific notation if you choose, using formats such as '1.23E1' or '1.23e1', '1.23D1' or '1.23d1'. The ESCape key will return you to the previous value. Warnings will be issued if you enter values outside of the current upper and lower bounds on that parameter.

3.2.4.2 Fit control check boxes



These check boxes control whether a parameter will be varied iteratively during the fit ("floated") or held constant at the value specified in the text box to the left ("fixed"). The text next to the check box will correspondingly read "Yes" or "No".

TIP: When a fit is having difficulty converging it is often helpful to temporarily fix one or more parameters so the fitter can get the other parameters into a reasonable range (giving a rough fit of the experimental data). Then return to the **Set/alter fitting parameters page (Section 3.2.4)** and turn on fitting of one more parameter, and re-fit, continuing this process until all parameters are actively fitted.

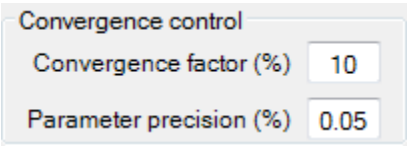
The boxes are inactive (grayed) and read "constrained" for **constrained parameters (Section 3.2.4.3)**.

3.2.4.3 Constrained parameters

Users cannot input values for those parameters whose values are being computed via constraints (see **Constrain M or s ratios among species (Section 3.2.3.7)**). The current values of those parameters are displayed with a light yellow background to indicate that these are computed quantities. The corresponding **Fit control check box (Section 3.2.4.2)** is also inactive (since the parameter is automatically being fitted) and reads "constrained" as a reminder of that condition.

Note that the displayed value will not change if you manually alter the parameter to which this parameter is linked---the constrained values are recomputed just prior to each fit iteration.

3.2.4.4 Convergence control



The **Convergence factor** and **Parameter precision factor** are numerical factors used in determining when a fit has converged. The criteria used to detect convergence are either

1. the fit iterations are producing insignificantly small changes in all the fitting parameters
2. the fit iterations are lowering the *sum of squared residuals* by an amount that is too small to be statistically significant.

Specifying a lower value for these factors will produce more iterations, but those iterations will be less and less likely to produce significant changes in parameter values. Specifying a larger value for them can help avoid wasted iterations, especially in situations where the parameters are highly correlated.

In any iterative fitting scheme the iterations can proceed indefinitely. However, if the fit is successful, at some point the sum of squared residuals will approach a minimum value and the changes in parameters between successive iteration will be very small, so further iterations are pointless. Therefore it is necessary to have a rule to define when the fit has converged.

The first criterion for convergence is based on a percentage precision on the parameters. Fitting continues until the changes in all parameters with each iteration are below this precision. By default this program will define the fit to have converged if the change in all parameters is less than 0.05%, but this can be altered from the **Set/alter fitting parameters page (Section 3.2.4)**.

The problem with using parameter precision as the only convergence criterion is that usually some of the parameters can be determined with much higher precision than others, and many many iterations can be spent trying to get high precision on a parameter whose standard error will be very large. In this program, it is almost always possible to determine the sedimentation coefficients with much higher precision than the diffusion coefficient (or molecular mass).

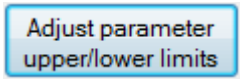
This program avoids this problem by also basing convergence on a second test, the size of the changes in the sum of squared residuals (and in general it is this criterion which halts the iterations first). Statistics tells us that a change in the value of a parameter will not be statistically significant unless that change increases the sum of squared residuals by a factor of approximately 1/N, where N is the number of data points being fitted. Thus once successive iterations of the fit produce a fractional drop

in the sum of squared residuals that is smaller than 1/N, further iterations will produce refinements in the parameters that are not statistically significant.

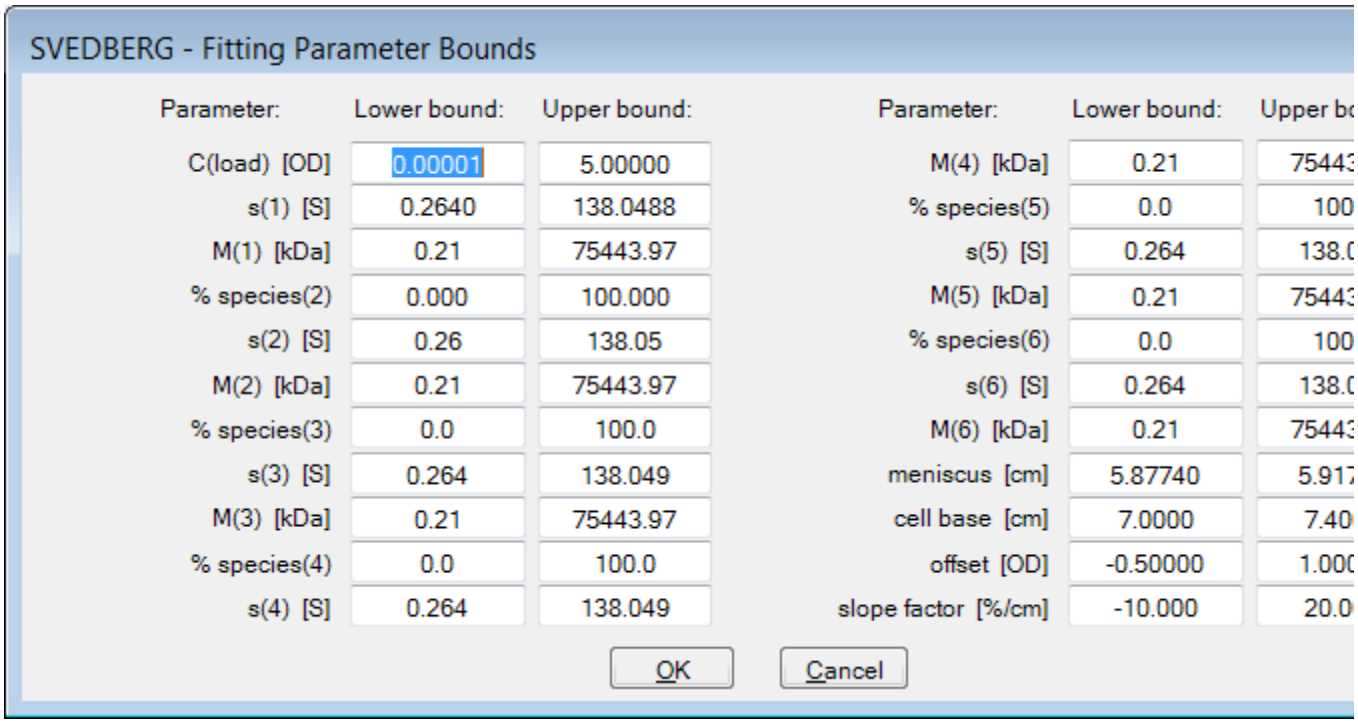
The actual implementation is that the fit will continue until the fractional drop in the sum of squared residuals is less than 1/N times the "convergence factor". Thus if the convergence factor is 0.01 (the default value), the fit continues until the changes are 100 times smaller than the smallest change that is likely to be statistically significant.

For multi-species fits, or in situations where the parameters are very highly correlated, it may be worthwhile to increase the values of the convergence factor or parameter precision factor to reduce computation time.

3.2.4.5 Adjust parameter upper/lower limits button



This button will bring up the dialog box below for altering the default upper and lower limits bounds) on parameter values.



- **What is the purpose of these limits?**

The intent of bounding the parameters is to prevent the fitter from straying into areas where the parameters would make no physical sense. Generally the fitter never needs to impose bounding, and the user will not need to alter the default values for the bounds, but in certain situations the program defaults may be inappropriate and could potentially prevent convergence.

- **How are the default limits established?**

1. Default limits for concentrations are calculated based on the data type (absorbance, interference or fluorescence). The lower limit is always slightly above zero signal.
    - If the **species 1 negative option (Section 3.2.3.1)** is selected the limits for species 1 are made negative and swapped in order.
2. Default limits for sedimentation coefficients are set based on estimates for the average sedimentation coefficient of the sample (automatically calculated from the loaded raw scans).
3. Default diffusion coefficient or molecular weight limits are then calculated based on those sedimentation coefficient limits and estimates of values for anhydrous spheres.
    - For example the maximum diffusion coefficient is assumed to be 3 times the anhydrous sphere value corresponding to the lowest sedimentation coefficient; the maximum *M* value is set as 10X that for an anhydrous sphere with the maximum sedimentation coefficient.)
4. The default limits for the meniscus position are +/- 0.02 cm from the current position (except when fitting synthetic boundary data, where the limits are 6.45 to 6.65 cm).
5. The default limits for the cell base are +/- 0.2 cm from the current position.
6. The default limits for the baseline offset are -0.5 to 1 OD, -10 to +40 fringes, or -100 to +100 FI units.
7. The default limits for the **slope factor (Section 3.2.4.6)** are -10% to +20%.

How to alter the limits

The user is presented with the dialog box shown above, containing a list of the current upper and lower bounds. These can be manually altered as desired.

Note that the alterations will apply only to the current **analysis window (Section 3.1.3)**---they do not change how the program computes default values.

The **OK button** will accept any alterations that have been made and exit the form; the **Cancel button** will leave the values unchanged.

What happens if a parameter goes outside the limits?

During each round of fitting the new parameters are tested against the limits. If any parameters will go out of bounds, the increments for all parameters are reduced proportionally using the **increment reduction factor (Section 3.2.5.6.1)** until none will go out of bounds. That procedure moves all parameters in the direction, and in the relative proportions, that should improve the quality of the fit. However this bounding procedure may prevent the 'good' parameters (the ones not exceeding the bounds) from ever reaching the values needed for an optimum fit, i.e. often when any parameter is going out of bounds the fit will never converge.

Therefore it is often useful to warn the user that parameters are trying to go out of bounds. This is done by automatically switching the Fit Monitor to the **Fitting parameters monitor tab (Section 3.2.5.6)** and displaying the value of that parameter with a bright yellow background.

3.2.4.6 Slope correction fitting parameter

This parameter was introduced in an attempt to compensate for the sloping plateau optical artifact. The value measures the slope as a percentage of the absorbance at a radial position of 6.5 cm, per cm of radial movement. For alternative strategies for dealing with a sloping plateau see **How to minimize the effects of a sloping plateau (Section 4.6)**.

In practice, the sloping plateau artifact typically produces a rise in absorbance across the plateau which is linear in radius and proportional to the total absorbance. No slope is seen near the meniscus.

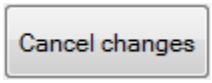
Thus the effect is similar to what would happen if the pathlength of the cell increased linearly with radius, and that is exactly how the theoretical curves are modified by this slope correction factor. That is, first the normal fitting function is calculated. Then at each radius, r, that value is multiplied by (slope parameter) $\times$ (r - 6.5) /

100.

Therefore when this parameter is in use, the loading concentration term becomes the absorbance as measured at a radial position of 6.5 cm.

When the sloping plateau problem is significantly present, for a plateau absorbance of 0.6-1 the value of this parameter is typically 5-8 %/cm; when the problem is severe it may exceed 10 %/cm. Under such conditions the values returned for diffusion coefficients or molecular masses are likely to have errors of 5% or more.

3.2.4.7 Cancel changes button



This button will return all the settings on the **Set/alter fitting parameters page (Section 3.2.4)** to the values they had when this page was first entered (it is an 'Undo' command).

3.2.5 Perform least-squares fit page

This analysis page controls and monitors the non-linear least-squares fitting process.

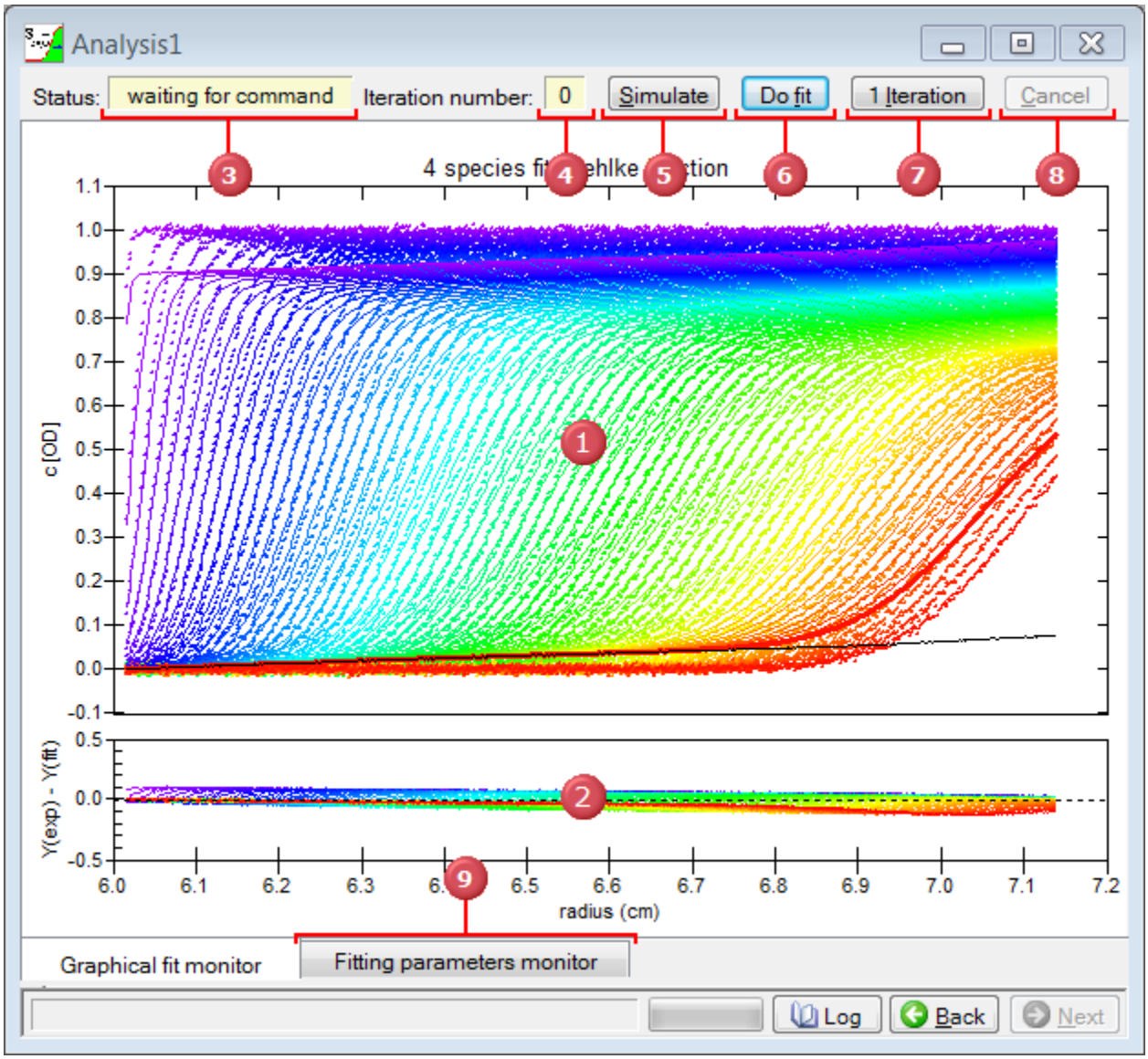
- In most cases you will simply push the **Do fit button (Section 3.2.5.1)** and watch the theoretical curve move toward the experimental data as the fit converges.
- However you are also provided options of proceeding one iteration at a time, or aborting the process, and of monitoring the actual parameter values and fit statistics rather than the graph.

This page is divided into two tabs:

1. the **Graphical fit monitor tab**, which will be shown and described below, and
2. the **Fitting parameters monitor tab (Section 3.2.5.6)**, which is described in detail in its own Help topic.

The **Do fit button (Section 3.2.5.1)**, **1 iteration button (Section 3.2.5.2)**, **Simulate button (Section 3.2.5.3)**, and **Cancel button (Section 3.2.5.4)** are common to both tabs, as are displays of the fit **Status (Section 3.2.5.5)** and the current iteration number.

Both tabs are always active, although the **Fitting parameters monitor tab (Section 3.2.5.6)** will be of more interest to advanced users.



The image above shows this analysis page immediately after entry from the **Set/alter fitting parameters page (Section 3.2.4)** for a four-species fit.

1 **Fit overlay and species plot** - The upper graph panel displays an overlay of the data to be fitted (displayed as individual dots) and the initial theoretical fit (continuous lines) based on the initial guesses for the parameters. In this case the initial guesses are not too close so there is a moderately large deviation between the fit and the data. This fit includes a fitted TIN, which is shown as the black line. (The heavier red line is due to a bug in the third-party graphing package).

2 **Residual plot** - The lower graph panel displays a *residual* plot (red line). The vertical scale of this plot will change as the fit proceeds and the residuals grow smaller.

3 **Status box** - The **Status box (Section 3.2.5.5)** will display status messages as the fit proceeds. In this case the program is waiting for the **Do fit button (Section 3.2.5.1)**, **1 iteration button (Section 3.2.5.2)**, **Simulate button (Section 3.2.5.3)**, or **Cancel button (Section 3.2.5.4)** to be pressed.

4 **Iteration counter** - The iteration counter simply displays the total number of iterations in this fitting

session (0 in this case since fitting has not started).

5 **Simulate button** - The **Simulate button (Section 3.2.5.3)** simply computes the theoretical curves based on the current parameter settings. The primary purpose is that this will enable movement to the **Graph fit results page (Section 3.2.7)**. Thus by manually modifying parameters on the **Set/alter fitting parameters page (Section 3.2.4)** and then using the **Simulate button (Section 3.2.5.3)** you can generate theoretical plots, residual plots, *etc.* for any desired set of parameters (not necessarily ones that give good fits).

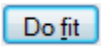
6 **Do fit button** - The **Do fit button (Section 3.2.5.1)** tells the fitter to proceed, without halting after each iteration. This is the normal mode, and thus this button is given the *focus* by default when you enter this page so you can initiate the fit by pressing the **Enter** key.

7 **1 iteration button** - The **1 iteration button (Section 3.2.5.2)** is used to manually step through the fitting procedure one iteration at a time. This can be a useful learning exercise, but generally this would only be used when you are having trouble obtaining a fit and want to carefully monitor what is happening, step-by-step.

8 **Cancel button** - The **Cancel button (Section 3.2.5.4)** will abort the current fit, stopping any calculations that are in progress. The **Back button (Section 3.1.3.1)** can then be used as needed to alter parameter guesses or the fitting model before trying to fit again.

9 **Fitting parameters monitor tab (Section 3.2.5.6)** - Select this tab to monitor the actual values of the parameters and the decrease of the *sum of squared residuals* as the fit progresses.

3.2.5.1 Perform fit - Do fit button

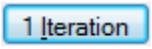


This button will allow the fit to proceed without pausing after each iteration.

This is the normal way to complete a fit. The fitting algorithm will iterate, seeking parameters that give the lowest *sum of squared residuals* (lowest χ^2 for weighted fits). The iterations will halt when the iterations no longer give improvement (the fit has converged). At that point the program will automatically move to the **Report fit results page (Section 3.2.6)** to display the results.

Once fitting is started, it can be aborted using the **Cancel button (Section 3.2.5.4)**.

3.2.5.2 Perform fit - 1 iteration button



This button will allow the fit to proceed through only one iteration, after which it will pause.

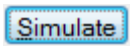
This procedure allows time for examination of trends in the parameter values, and whether the initial guesses for the parameters are reasonable. This may be a good choice for the initial stages of a difficult fit. If the parameters are moving in a direction that makes no physical sense, you can go back to the **Set/alter fitting parameters page (Section 3.2.4)** and perhaps try holding one or more parameters constant until the other parameters get into a better range, or perhaps try using **parameter constraints (Section 3.2.3.7)**.

At any point the fit may be continued without pausing by using the **Do Fit button (Section 3.2.5.1)**.

If the requested single iteration produces convergence the program will automatically move to the **Report fit**

results page (Section 3.2.6) to display the results.

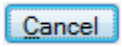
3.2.5.3 Perform fit - Simulate button



The **Simulate button** forces the program to calculate the theoretical fit and the *residual*s using the current parameters and model, as though a fit has been completed. Those calculations will enable the display of the graphs on the **Graph fit results page (Section 3.2.7)**.

By manually altering the parameters via the **Set/alter fitting parameters page (Section 3.2.4)** and then using the **Simulate button** you can generate theoretical curves for specific values, graph those results, and if desired even export the theoretical data.

3.2.5.4 Perform fit - Cancel button



This will abort the current fit or single iteration, stopping any calculations that are in progress.

The **Back button (Section 3.1.3.1)** can then be used as needed to go to the **Set/alter fitting parameters page (Section 3.2.4)** to alter parameter guesses or change which parameters are being fitted, or to go to the **Select fitting model page (Section 3.2.3)** to change the fitting model before trying to fit again.

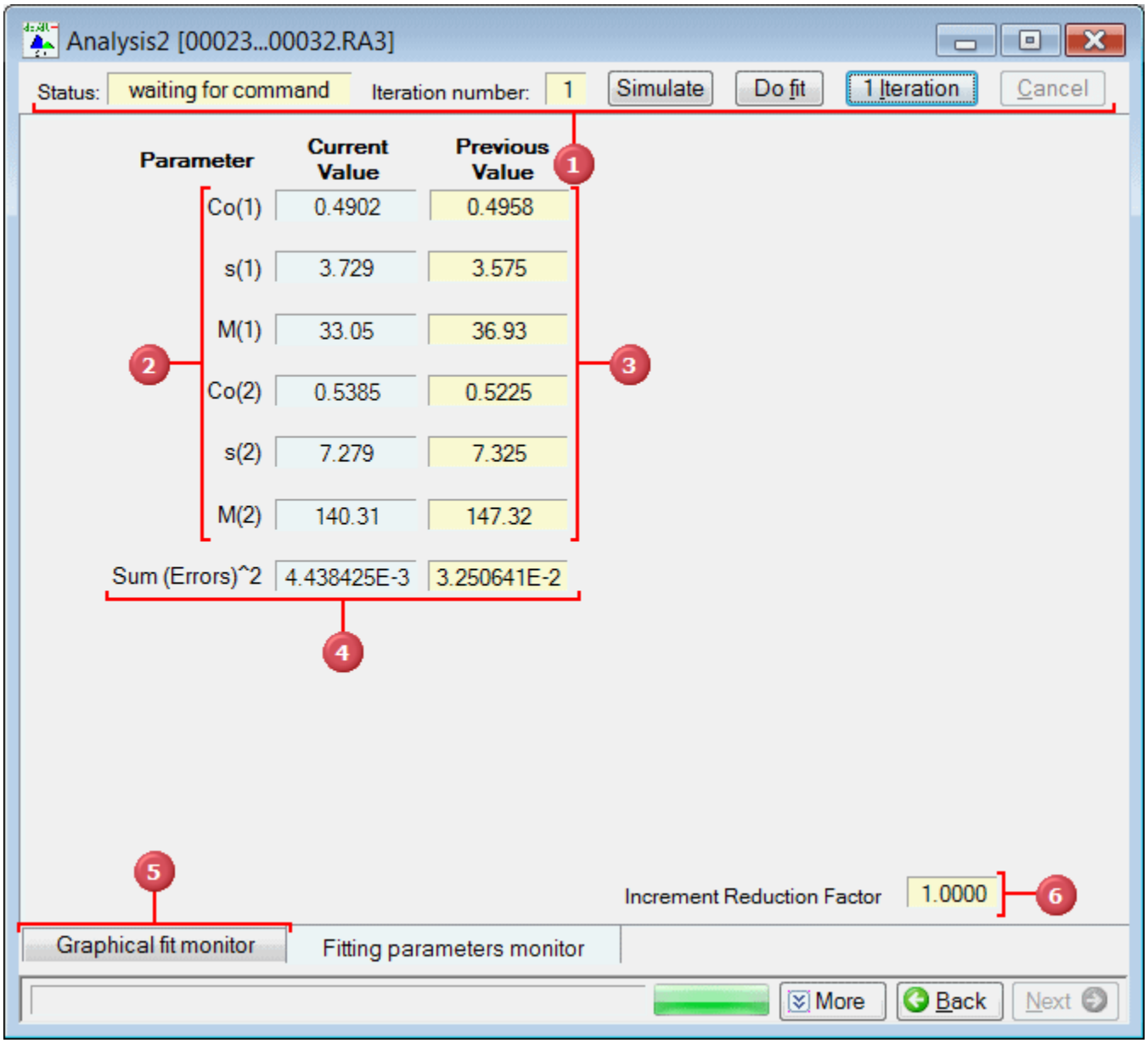
3.2.5.5 Perform fit - Status

This **status box** shows status messages as the fit progresses:

- **Waiting for input** indicates the program is halted waiting for the **Do Fit (Section 3.2.5.1)**, **1 Iteration (Section 3.2.5.2)**, or **Simulate (Section 3.2.5.3)** button
- **Calculating sums** indicates the program is looping through the data points collecting the *sum of squared residuals* and partial derivatives that are needed to calculate improved parameters
- **Inverting matrix** indicates the matrix inversion needed for calculating improved parameters is being done
- **Adjusting increment** indicates that the new parameters are being tested to be sure the *sum of squared residuals* is improved. During this time the **increment reduction factor (Section 3.2.5.6.1)** may change.
- **Converged, last sum** indicates that convergence has been obtained and a final set of sums needed for calculating parameter errors is being accumulated

3.2.5.6 Fitting parameters monitor tab

This tab is used to monitor the actual values of the fitting parameters as they are updated during the fit.



The image above shows the analysis page after the first iteration of a fit, which was initiated via the **1 iteration button (Section 3.2.5.2)**.

1 **Fit control buttons and displays** - These controls are shared with the **Graphical fit monitor tab**. Follow **this link (Section 3.2.5)** for descriptions and links to further details.

2 **Current value column -** This column shows the parameter values for the current iteration of the fit. Comparisons to the **Previous value** column to the right will show whether this parameter is changing rapidly or converging to a definite value.

- Before the first iteration is completed these values will correspond to those entered or computed as the initial guesses
- If a parameter tries to go outside the current upper or lower limits the background color of its text box will flash momentarily bright yellow. It will stay yellow once that parameter reaches the limit. (Note that when parameters go out of bounds the program will switch from the **Graphical fit monitor tab** to show this **Fitting parameters monitor tab** if needed.)

3 **Previous value column** - This column shows the parameter values from the previous iteration of the fit.

- Before the first iteration is completed this column will be empty.

4 **Sum of squared residuals -** This row displays the *sum of squared residuals* calculated for the current and previous iteration.

5 **Graphical fit monitor tab (Section 3.2.5)** - Select this tab to monitor fit overlay and residual plots as the fit progresses.

6 **Increment reduction factor** - The **increment reduction factor (Section 3.2.5.6.1)** is used within a "modified Gauss-Newton fitter" to try to ensure that each iteration actually lowers the variance. When the value falls below 1 that indicates the fitter is having difficulty finding improved parameters. It is normal, however, for this value to fall just before the fit converges because further improvement is not possible.

3.2.5.6.1 Increment reduction factor

The **increment reduction factor** is a numerical multiplier used within a so-called "modified Gauss-Newton" fitter to improve the rate of convergence [**Reference 2 (Section 7.2)**]. When this value falls below the nominal value of 1, this is an indication that the fitter either

1. is having some difficulty finding a good fit (there are too many species or very poor initial guesses), or
2. it is so near to convergence that it simply cannot find better parameter values

How it works

Let {p1, p2, ...} be the set of current fitting parameters. Then at each iteration the Gauss-Newton method calculates a set of increments to those parameters, {δp1, δp2, ...}, that will define a new set of improved parameters, {p1+ δp1, p2+ δp2, ...}, which should give a lower *variance*.

In the modified method these increments are each multiplied by a scalar factor, α, the "increment reduction factor", and hence the new parameters will be {p1 + α x δp1, p2 + α x δp2, ...}.

This modification is used because when the fit is far from the optimal parameters the Gauss-Newton method tends to pick a reasonable direction for the increment vector, but not necessarily the right magnitude. Thus the idea is to test how the variance changes along this direction, and pick a value of α that will, at least, reduce the variance.

When each fit begins α is initially set to 1. At each iteration, the variance is first calculated for a new set of parameters, using an initial value of α based on that from the preceding iteration (see rule 1 below). If these new parameters give a lower variance than that for the previous parameter set, these parameters are accepted and the next iteration is begun.

If instead the new parameters give a higher variance, then the value of α is reduced by factors of two until finally a lower variance is obtained (unless the value of α falls below 0.001, which triggers exception 1 below). Once this lower variance is obtained, this variance together with those for the preceding values of α are used to fit a parabolic function of variance *versus* α and thereby to predict what value for α will give the minimum variance. This predicted value for α is then used to calculate the parameter set for the next iteration of fitting.

Although these procedures are intended to guarantee that the variance will drop at each and every iteration, occasionally that will not be true.

Rule 1 adjusts the starting value for α when α from the preceding iteration is small. In this circumstance it is hoped that this previous iteration succeeded in moving into a region of parameter space where convergence will be better, and therefore α can be increased back toward 1. Further, since it is much faster to calculate variances in order to test values for α than it is to carry out a full iteration, it is usually worthwhile to try

keeping α large in order to move more rapidly to convergence. Therefore the rule applied is that when α from the preceding round is < 0.005 it is set to 0.02; for 0.005 < α < 0.25 it is multiplied by 4, and for 0.25 < α < 1 it is set to 1.

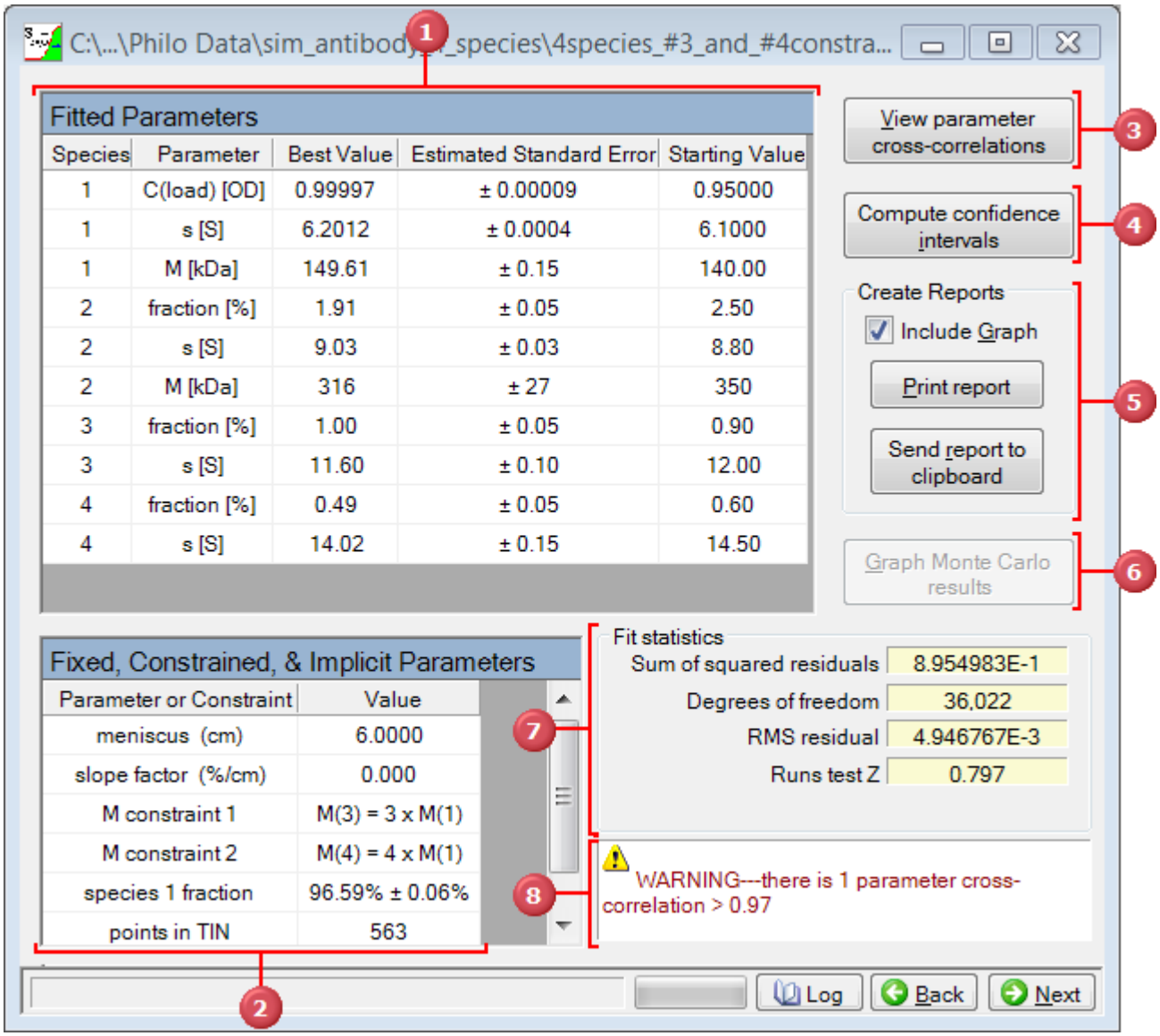
Exception 1 occurs when the sequential reductions of α during the attempt to find a lower variance reduce α below 0.001. This means the fitter is having major trouble! That may be because too many species are being included in the fitting model, or because one or more of those species are poorly resolved or in very low abundance. In that circumstance the program tries making α negative (taking the parameters in the opposite direction, since even small positive values of α just make things worse). If a negative α lowers the variance, that value is accepted; otherwise the sign of α is inverted again and that positive value accepted. If a negative value of α is used 5 iterations in a row, the program assumes convergence has been reached.

3.2.6 Report fit results page

The **Report fit results** page is displayed automatically upon completion of a fit, or you can navigate to it whenever fit results are available.

This page displays the best fit parameter values, the values for parameters that were held constant or constrained, the fit statistics, and buttons for printing a report.

Each area or control on the image below is a clickable *hotspot* that will lead to further information, or just use the numbered list below the image.


C:\...\Philo Data\sim_antibody\3_species\4species_#3_and_#4constra...
Fitted Parameters
Species | Parameter | Best Value | Estimated Standard Error | Starting Value
1 | C(load) [OD] | 0.99997 | ± 0.00009 | 0.95000
1 | s [S] | 6.2012 | ± 0.0004 | 6.1000
1 | M [kDa] | 149.61 | ± 0.15 | 140.00
2 | fraction [%] | 1.91 | ± 0.05 | 2.50
2 | s [S] | 9.03 | ± 0.03 | 8.80
2 | M [kDa] | 316 | ± 27 | 350
3 | fraction [%] | 1.00 | ± 0.05 | 0.90
3 | s [S] | 11.60 | ± 0.10 | 12.00
4 | fraction [%] | 0.49 | ± 0.05 | 0.60
4 | s [S] | 14.02 | ± 0.15 | 14.50
View parameter cross-correlations
Compute confidence intervals
Create Reports
Include Graph
Print report
Send report to clipboard
Graph Monte Carlo results
Fixed, Constrained, & Implicit Parameters
Parameter or Constraint | Value
meniscus (cm) | 6.0000
slope factor (%/cm) | 0.000
M constraint 1 | M(3) = 3 x M(1)
M constraint 2 | M(4) = 4 x M(1)
species 1 fraction | 96.59% ± 0.06%
points in TIN | 563
Fit statistics
Sum of squared residuals | 8.954983E-1
Degrees of freedom | 36,022
RMS residual | 4.946767E-3
Runs test Z | 0.797
WARNING---there is 1 parameter cross-correlation > 0.97
Log
Back
Next
1
2
3
4
5
6
7
8

The image above shows the results from a 4-species fit where the masses of species 3 and 4 have been constrained as trimer and tetramer of species 1.

1 **Fitted parameters (Section 3.2.6.1)** - This table lists the best fit parameters, the estimated uncertainty for that parameter (the standard error or true confidence limits after those have been computed), and the initial value for that parameter at the start of the fit procedure.

2 **Fixed, constrained & implicit parameters (Section 3.2.6.2)** - This table lists the values for parameters that were held constant during the fit, any parameter constraints that were imposed. It also displays implicit values such as the fraction of species 1 (when the **Fit total concentration and species fractions option (Section 3.2.3.1)** applies, as in this example), the total loading concentration (when there are multiple species and species fractions are not used), and the total number of data points used to create the TIN (as in this example).

3 **View parameter cross correlations (Section 3.2.6.3) -** This button will bring up a dialog box displaying the parameter cross-correlation matrix. See **What does high parameter cross-correlation mean? (Section 5.1.6)**.

4 **Compute confidence intervals (Section 3.2.6.4)** - This button initiates a procedure to calculate more rigorous estimates of the uncertainties in the parameter values. Generally you will only do this when you are satisfied this is your final result. (The button is disabled in the example above because this computation has already been completed).

5 **Print report buttons (Section 3.2.6.5)** - These buttons will print a report which documents every aspect of your analysis. The lower button prints a report that also includes a summary graph showing an overlay of the data and the fit as well as a residual plot.

6 **Graph bootstrap or Monte Carlo results (Section 3.2.6.8)** - This button is active whenever confidence intervals have been computed using either the bootstrap or Monte Carlo methods (the name on the button changes to whichever method was last used). It will open a dialog box that generates graphs of the parameter distributions (either histograms of a single parameter, or two-parameter correlation plots).

7 **Fit statistics (Section 3.2.6.6)** - This area displays some of the statistical parameters characterizing the fit.

8 **High cross-correlation warning (Section 3.2.6.7)** - This warning panel is only displayed when there are one or more pairs of parameters that are highly correlated (there is one in this example). This is a warning that some of the parameters may be poorly determined, and you may want to consider fixing or constraining some of the parameters.

3.2.6.1 Report fit results - Fitted parameters

Fitted Parameters

Species	Parameter	Best Value	Estimated Standard Error	Starting Value
1	C(load) [OD]	0.88461	± 0.00009	0.88459
1	s [S]	6.3901	± 0.0008	6.3948
1	M [kDa]	146.63	± 0.18	146.36
2	fraction [%]	3.44	± 0.07	2.78
2	s [S]	9.40	± 0.04	9.39
2	M [kDa]	125	± 6	293
3	fraction [%]	0	± 9	0
3	s [S]	14	± 7	12
4	fraction [%]	0.1	± 1.6	0.1
4	s [S]	15.7	± 2.4	15.6
-	meniscus [cm]	5.90737	± 0.00004	5.90719

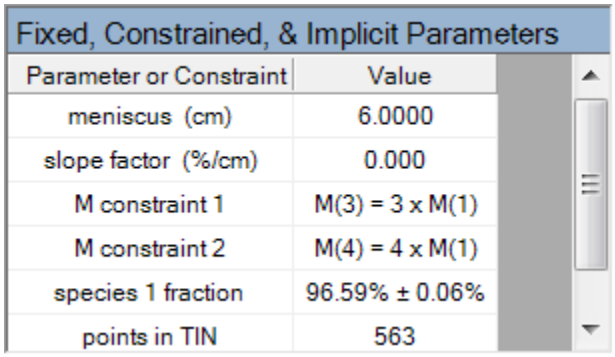
This area of the **Report fit results page (Section 3.2.6)** displays:

1. the returned best-fit values for the fitted parameters.
    - Note that the units applicable to the fitting parameters are shown; these will depend on other program choices and options, such as whether you are fitting concentrations or species fractions.
2. an estimate of the error (uncertainty) for each parameter
    - When the fit is first completed that error estimate will be a simple estimate of the standard error based on the covariance matrix, as in the example above

- After the **(Section 3.2.6.4)Compute confidence intervals button (Section 3.2.6.4)** has been used, this area will instead show the true confidence intervals, with the confidence level shown in the column heading.

3. The starting value for each parameter (the initial guess at the start of that fitting session).
   - This is provided primarily to document the initial conditions of the fit.
   - Note that fits saved from versions 1-6 of this program did not store that information, so this column will not be displayed.

3.2.6.2 Report fit results - Fixed, constrained, & implicit parameters

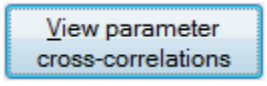


This table lists:

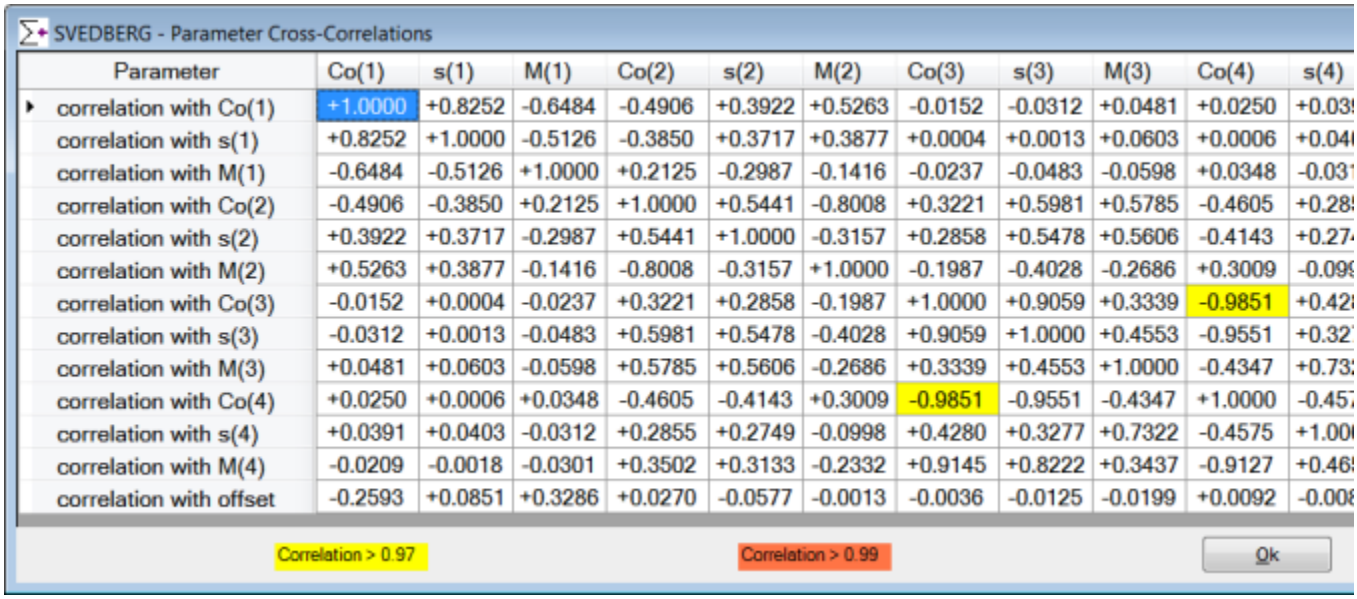
1. The values for parameters that were held constant during the fit (parameters whose values were not optimized during the fit). In this example the position of the meniscus has been fixed at 6.000 cm (this is simulated data) and the slope factor has been held fixed at zero.
2. Equations summarizing any parameter constraints that were imposed. In this example the mass of species 3 is constrained to be 3 times the mass of species 1 (species 3 is a trimer of species 1), and species 4 is constrained to be a tetramer of species 1.
3. Implicit values such as the fraction of species 1 (when the **Fit total concentration and species fractions option (Section 3.2.3.1)** applies, as in this example), the total loading concentration (when there are multiple species and species fractions are not used), and the total number of data points used to create the TIN (as in this example).

   - Note that if confidence limits from the bootstrap or Monte Carlo methods are available then true confidence limits for the total loading concentration will be calculated from the actual distribution and reported.
   - For the species 1 fraction, if confidence limits from the bootstrap or Monte Carlo methods are available then true confidence limits for the species 1 fraction will be calculated from the actual distribution and reported. Otherwise a precision *estimate* (a +/- range) is made by assuming the errors in each species fraction (as displayed in the **Fitted parameters (Section 3.2.6.1)** table)are random and independent (which is likely not actually true).

3.2.6.3 Report fit results - View parameter cross-correlations button



This button will bring up a window showing the cross-correlation matrix. Parameters with 'high' or 'very high' correlations will be flagged with yellow or red backgrounds. The graphic below shows this window.

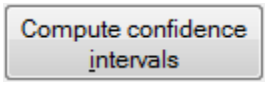


Parameter	Co(1)	s(1)	M(1)	Co(2)	s(2)	M(2)	Co(3)	s(3)	M(3)	Co(4)	s(4)
correlation with Co(1)	+1.0000	+0.8252	-0.6484	-0.4906	+0.3922	+0.5263	-0.0152	-0.0312	+0.0481	+0.0250	+0.039
correlation with s(1)	+0.8252	+1.0000	-0.5126	-0.3850	+0.3717	+0.3877	+0.0004	+0.0013	+0.0603	+0.0006	+0.040
correlation with M(1)	-0.6484	-0.5126	+1.0000	+0.2125	-0.2987	-0.1416	-0.0237	-0.0483	-0.0598	+0.0348	-0.031
correlation with Co(2)	-0.4906	-0.3850	+0.2125	+1.0000	+0.5441	-0.8008	+0.3221	+0.5981	+0.5785	-0.4605	+0.285
correlation with s(2)	+0.3922	+0.3717	-0.2987	+0.5441	+1.0000	-0.3157	+0.2858	+0.5478	+0.5606	-0.4143	+0.274
correlation with M(2)	+0.5263	+0.3877	-0.1416	-0.8008	-0.3157	+1.0000	-0.1987	-0.4028	-0.2686	+0.3009	-0.099
correlation with Co(3)	-0.0152	+0.0004	-0.0237	+0.3221	+0.2858	-0.1987	+1.0000	+0.9059	+0.3339	-0.9851	+0.428
correlation with s(3)	-0.0312	+0.0013	-0.0483	+0.5981	+0.5478	-0.4028	+0.9059	+1.0000	+0.4553	-0.9551	+0.327
correlation with M(3)	+0.0481	+0.0603	-0.0598	+0.5785	+0.5606	-0.2686	+0.3339	+0.4553	+1.0000	-0.4347	+0.732
correlation with Co(4)	+0.0250	+0.0006	+0.0348	-0.4605	-0.4143	+0.3009	-0.9851	-0.9551	-0.4347	+1.0000	-0.457
correlation with s(4)	+0.0391	+0.0403	-0.0312	+0.2855	+0.2749	-0.0998	+0.4280	+0.3277	+0.7322	-0.4575	+1.000
correlation with M(4)	-0.0209	-0.0018	-0.0301	+0.3502	+0.3133	-0.2332	+0.9145	+0.8222	+0.3437	-0.9127	+0.465
correlation with offset	-0.2593	+0.0851	+0.3286	+0.0270	-0.0577	-0.0013	-0.0036	-0.0125	-0.0199	+0.0092	-0.008

In this particular example there is a high (in absolute value) negative correlation between the concentration parameters for species 3 and 4. That means the increase in variance that results when the concentration of species 3 is increased can be almost completely removed by decreasing the concentration of species 4. The negative correlation between the sedimentation coefficient of species 3 and the concentration of species 4 is also quite strong (-0.9551), as is the positive one between the concentration of species 3 and the mass of species 4 (+0.9145), although these are not quite strong enough to flag in this case.

For this sample species 3 and 4 (trimer and tetramer of an antibody) are present at levels of only 1% and 0.5%, and there is simply insufficient information in the data to determine the concentration, sedimentation coefficient, and mass accurately for species 4. This is a case where it would be better to constrain the mass of species 4, which will improve the precision of the estimates for the concentration and sedimentation coefficient of not only species 3 but also species 1 and 2.

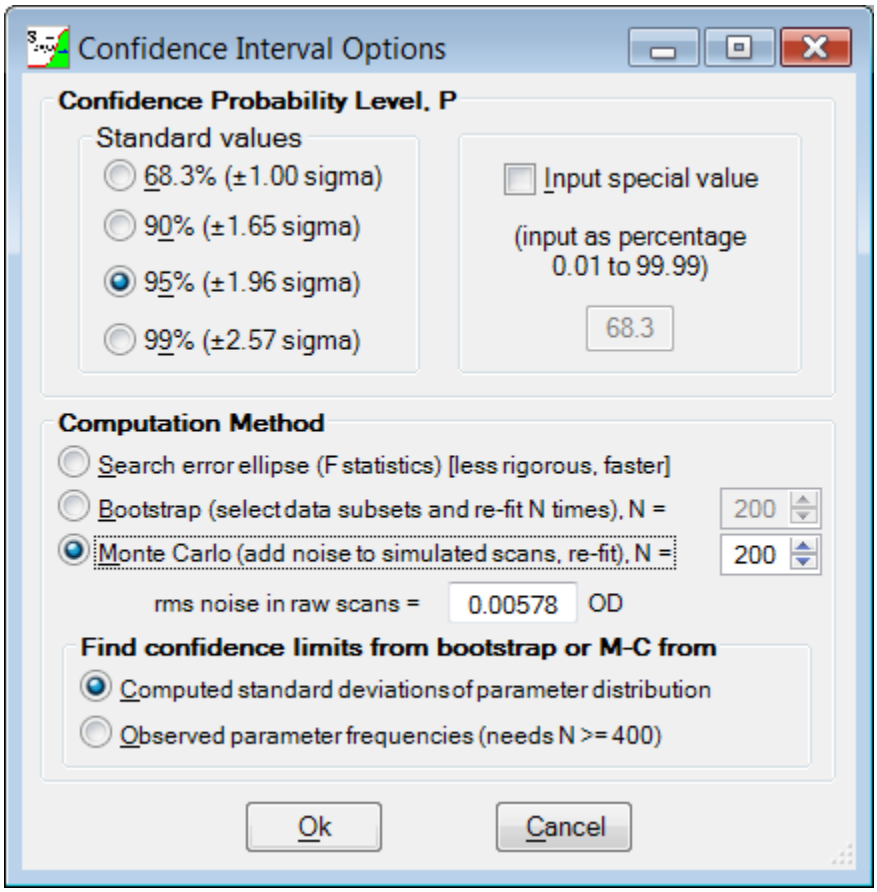
3.2.6.4 Report fit results - Compute confidence intervals button



This button will initiate computation of confidence intervals for the parameters, a much more statistically rigorous way to express the uncertainties in the best fit parameters (but sometimes quite time consuming to calculate).

TIP: Don't bother to compute confidence intervals until you are sure this is your final fit.

You are first presented with a dialog box for setting the desired confidence level, and selecting a computation method, as shown below. The default level of confidence that will be shown in this dialog (68.3%, 95%, *etc.*) can be altered via the **Options dialog (Section 3.1.1.4.3)**. The Ok button initiates the confidence level calculations.



The confidence limits are often not symmetric around the best fit value, and thus true confidence limits usually cannot be reported as a ± value.

Selecting a computation method:

Three radio buttons allow the user to select a preferred mathematical approach for computing the confidence intervals. These are:

Search error ellipse: This method moves each parameter away from its best-fit value until the *chi squared* increases to a limit that is bases on *F* statistics (it evaluates the shape of the error surface near the minimum). The implementation is based on **reference 2 (Section 7.2)**. The search of parameter space is made along axes of a hyper-ellipse that represent eigenvectors of the covariance matrix. The calculation of the critical *chi squared* level is based on equation 35 in Johnson & Straume (1994), **reference 11 (Section 7.11)**.

This approach is relatively fast to calculate. It's major drawback is that it is based on certain statistical models and assumptions, including an assumption that all the data points are statistically independent. Further, there is some disagreement among different sources about exactly how the critical *chi squared* level should be calculated.

Bootstrap with replacement: In this method data points are selected randomly from the original fitted data set to make a new data set with the same number of points (but due to the random selection some data points will be selected more than once). This new data set is then fitted, those parameters are stored, and the process is repeated many times to build up a distribution of parameter values. The confidence limits are then calculated from that parameter distribution (based on options as described below).

The up/down spinner next to this radio button sets the number of bootstrap rounds (default 200).

TIP: When the number of data points and species being fitted is high, the time to compute 200 rounds

of bootstrap or Monte Carlo may become excessive. As few as 50 rounds gives a reasonably good estimate of the standard deviation of the parameter distribution.

See **About the bootstrap method (Section 6.6)** for more details.

Monte Carlo: In the Monte Carlo method random noise is added to a set of theoretical scans that are calculated based on the current best-fit parameters. This new data set is then fitted, those parameters are stored, and the process is repeated many times to build up a distribution of parameter values. The confidence limits are then calculated from that parameter distribution (based on options as described below).

Not that because the Monte Carlo method assumes the noise is constant across the scans and both early and late in the run, which is usually not true for fluorescence scans, the bootstrap method may be a better choice for fluorescence scans (and this same issue applies to a lesser extent for absorbance scans).

The up/down spinner next to this radio button sets the number of Monte Carlo rounds (default 200).

See **About the Monte Carlo method (Section 6.7)** for more details.

The rms noise level to add to the raw scans can be entered manually in the text box below this radio button. If you are fitting simulated data you should enter the actual value used in the simulation. When you first bring up this dialog after a new fit an estimate of the noise in the raw data is made based on the *rms residual* for the fit and automatically entered into this text box. When scan differencing (delta-*c* mode) is being used, the noise level in each scan is estimated as the *rms residual* divided by the square root of 2.

Finding the confidence limits from the bootstrap or Monte Carlo results

Once a table of parameter values is built from the *N* rounds of bootstrap or Monte Carlo, there are two choices for calculating the parameter confidence intervals:

Computed standard deviations of the parameter distribution: With this choice first for each parameter the mean and standard deviation of the set of *N* values for that parameter is calculated, and then the confidence limit is found assuming that distribution is random (a Gaussian distribution). This option is generally the best choice when the number of rounds is 200 or less.

Because in reality for non-linear fits the parameter distribution may be asymmetric around the mean, two standard deviations are calculated, one for those parameters that exceed the mean value, and another for those that are less than the mean value. Those two standard deviations are then used to calculate an upper and lower confidence limit, respectively, and those values are not in general symmetric about the mean.

- It is also sometimes true that the mean parameter value from the bootstrap or Monte Carlo rounds is significantly shifted away from the corresponding best-fit parameter. When this is true (specifically when the shift is greater than 20% of the standard deviation) the confidence limits are also shifted (to make the mean match the best-fit value).

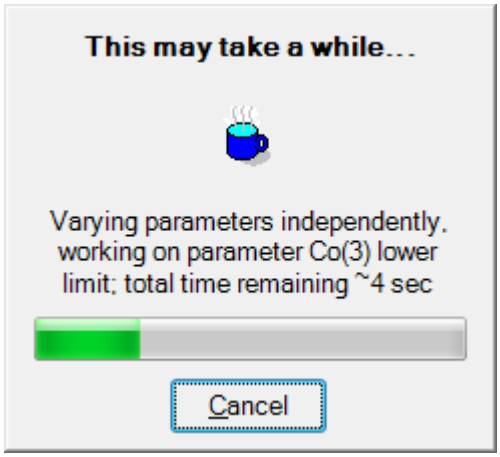
Observed parameter frequencies: With this choice no statistical assumptions are made about the shape of the parameter distribution. Instead, the *N* parameter values are first sorted and then to find the confidence limit the list is searched inward from the beginning or end (the lowest or highest observed values) until the appropriate confidence frequency has been exceeded. For example, for 95% confidence probability and *N* = 1000, the confidence limits correspond to frequencies of 2.5% (giving 5% total for the upper and lower wings of the distribution), which is 25 data points out of 1000. Therefore the confidence limits are the 25th and 975th values in the sorted list.

The drawback of this approach is to accurately find the desired confidence limit a large number of rounds, *N*, is required. If, for example, *N* is only 100, then it is not possible to distinguish a parameter frequency of 2.5% (the desired value for the lower wing for 95% confidence) from a frequency of 2% or 3%. That is, if you wanted the confidence probability to be accurate to 0.1%, it is necessary to have *N* >= 2000.

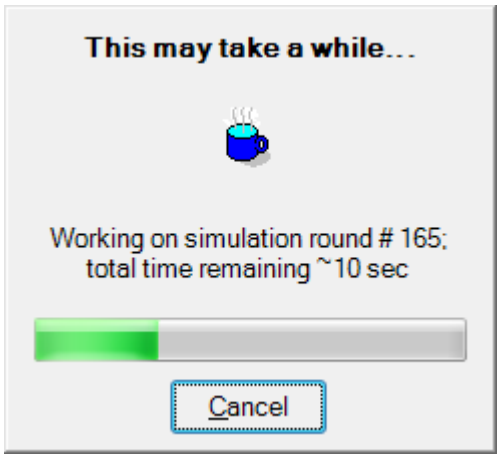
What happens after I have selected these options?

The **Ok button** will proceed to actually computing the confidence, and **Cancel** will abort that process. During the subsequent computations a small dialog box is displayed.

If you have selected the **Search error ellipse** option you will see a dialog box like that below, showing which confidence limit is currently being evaluated, an estimate of remaining time, and a progress bar.


This may take a while...
Varying parameters independently, working on parameter Co(3) lower limit; total time remaining ~4 sec
Cancel

For either the **Bootstrap** or **Monte Carlo** options the dialog box shown below which indicates the current simulation round, an estimate of remaining time, and a progress bar.

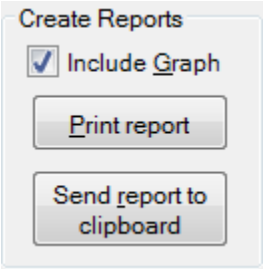


The **Cancel** button on these dialog boxes will abort the calculations (leaving the previous confidence limits intact if they exist).

What if I want to re-compute the confidence limits with different choices?

You can always change your selection and re-compute with a different confidence level or different computation method simply by pushing the **Compute confidence button** again.

3.2.6.5 Report fit results - Print reports



Either of the two buttons within this group box will generate a complete, nicely-formatted report defining what data were fitted, how they were fitted, the best fit parameters, fit statistics, etc. That report is then either printed or copied to the Clipboard (in rich text format) so that it can be pasted into a word processor or electronic lab notebook.

This report is intended to document every aspect of your analysis. All user choices and program defaults that affect the calculations of the data or the fitting procedure are listed in the report.

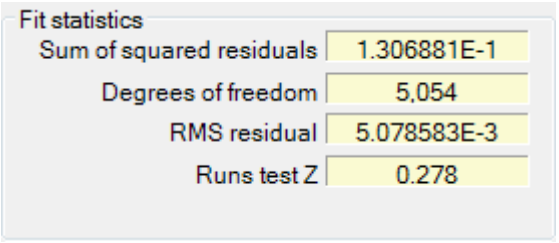
The **Include Graph checkbox** governs whether a summary graph showing an overlay of the data and the fit, with a residual plot underneath (as shown on the Graphical fit monitor tab of the **Perform least-squares fit page (Section 3.2.5)**) will be included at the end of the report.

If the **Print report** button is selected then a standard Windows printer dialog box will open allowing selection of the printer to be used for the report (this will by default be the printer set as the default report printer via the **Options dialog (Section 3.1.1.4.3)**. The graph will be in color if that printer is a color printer.

The size of the text font used for reports (whether printed or sent to the Clipboard) may be altered using the **Options dialog (Section 3.1.1.4.3)**.

Caution: Although the **Send report to Clipboard** button does not actually generated printed output, when a graph is to be included the program must have access to a valid printer in order to size the graph correctly relative to the text. If the default report printer is not valid either because it is not currently connected or no printer with that name exists, you will get a message asking for a valid printer (which could be a PDF printer or other virtual printer if you aren't currently connected to a real printer). Also note that the graph produced by this rich-text copy operation will be in color whether or not this printer is actually capable of color.

3.2.6.6 Report fit results - Fit statistics

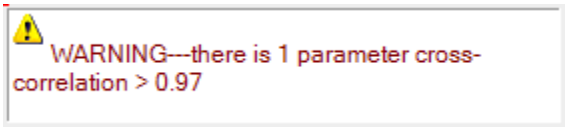


This area of the **Report fit results page (Section 3.2.6)** will report a statistical summary of the fit, including:

1. The sum of the squared residuals
2. The *degrees of freedom*. Note that the degrees of freedom correctly accounts for the fitting parameters needed for fitting the TIN and RIN (when those are used).
3. The *rms residual*
4. The *runs test Z* value (Z score). This is a measure of whether the residuals show excessive 'runs' (residuals which are all positive or all negative over adjacent regions). Long positive and negative runs are an indication that the residuals show systematic patterns (they are not random). A perfect fit has a Z score of

1 or less.

3.2.6.7 Report fit results - High cross-correlation warning



This area of the **Report fit results page (Section 3.2.6)** is used to display a warning whenever some of the parameters are highly cross-correlated. See **What does high parameter cross-correlation mean? (Section 5.1.6)**.

There are two levels of warning, one for "high" correlation and one for "very high" correlation. The numbers of correlations exceeding each level is listed, unless there are none at that level.

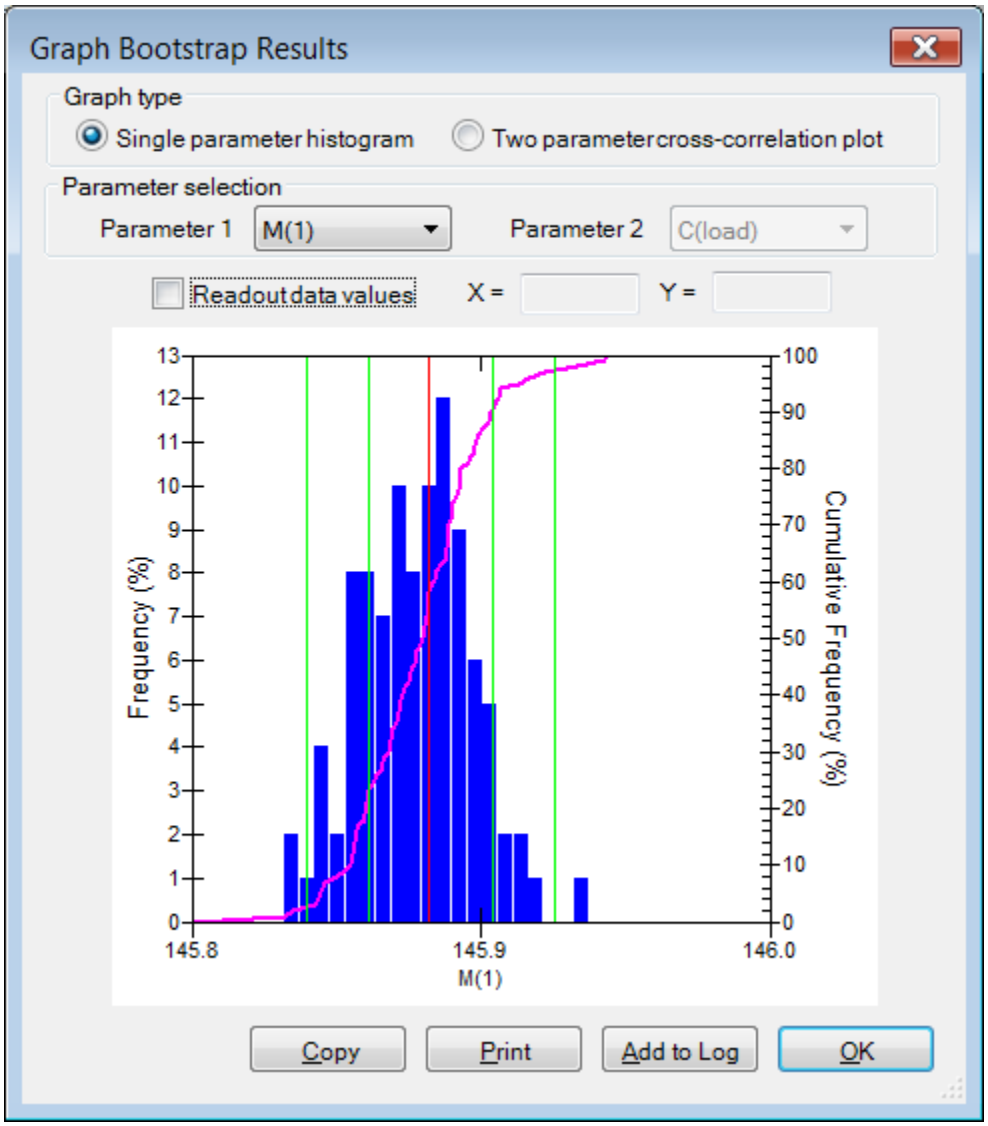
- The default levels that trigger those warnings are 0.97 and 0.99, respectively, but these levels can be set within the Fitting Options tab of the **(Section 3.1.1.4.3)Options dialog (Section 3.1.1.4.3)**.
- These warnings will also appear in **printed reports (Section 3.2.6.5)**.

The actual cross-correlation matrix can be viewed by using the **view parameter cross-correlations button (Section 3.2.6.3)** on this page.

3.2.6.8 Report fit results - Graph bootstrap or Monte Carlo results

These buttons become active whenever confidence limits have been calculated using either the Monte Carlo or bootstrap methods, respectively (the button name changes to match the method that was last used). Pressing the button will bring up a dialog box that displays graphs of the parameter distributions (using the table of parameter values from each round of bootstrap or Monte Carlo).

Initially this dialog box will display a histogram of the mass values for species 1 (assuming that parameter was fitted), like that shown below. The vertical red line marks the best-fit value. The vertical green lines on the histogram show the ±1σ and ±2σ limits (as calculated separately for the values above and below the mean). The magenta curve shows the cumulative frequency (the area under the distribution curve).

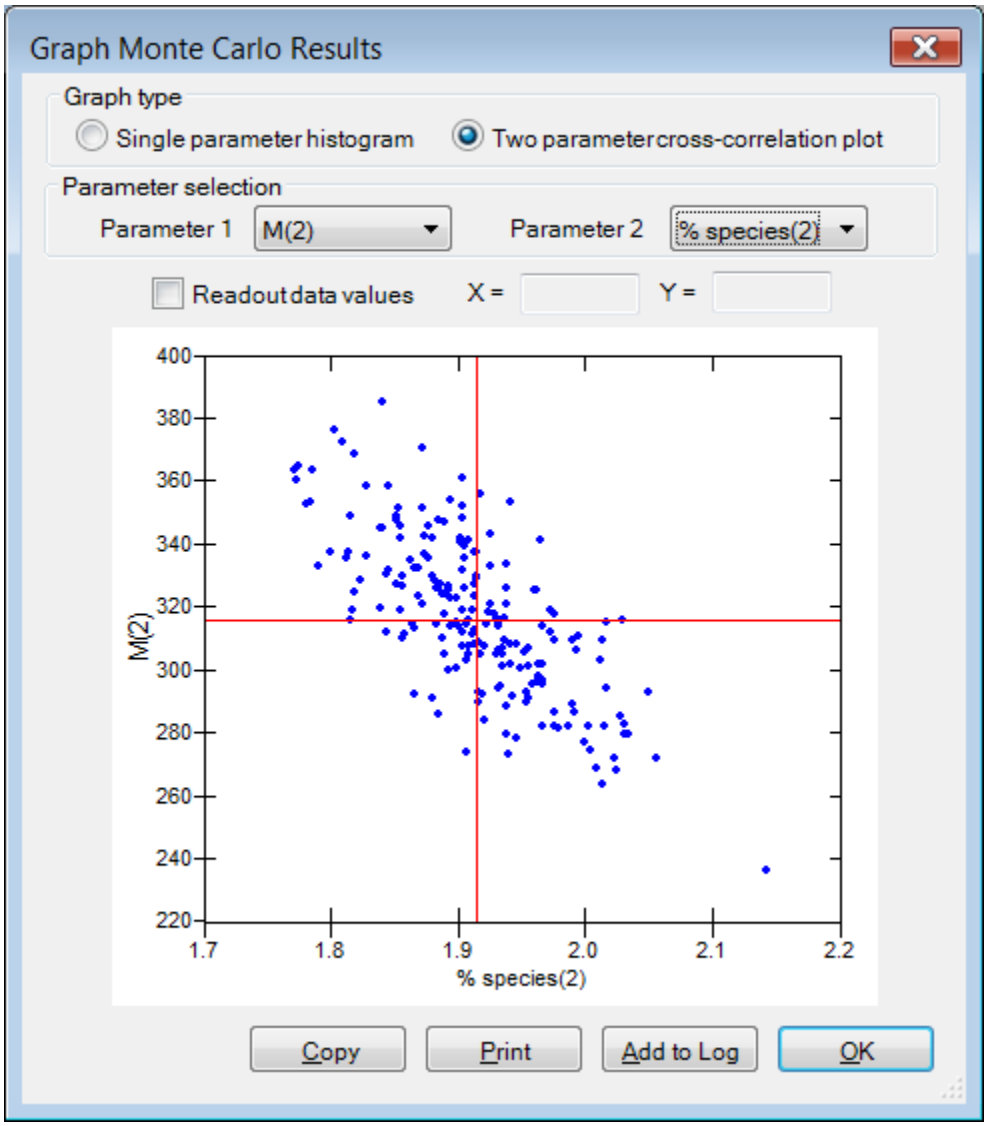


Corresponding graphs for the other fitting parameters can be obtained using the drop-down list labeled 'Parameter 1'.

TIP: The number of bins used for generating the histogram can be altered by clicking on any of the blue bars with the mouse. Using fewer bars will tend to average out the statistical fluctuations, but at a cost of fewer positions to define the true shape of the distribution. The program tries to set a number of bins that keeps the statistical fluctuations reasonably low while still providing reasonable definition of the shape of the distribution.

Parameter cross-correlation graphs

Clicking the **Two parameter cross-correlation plot radio button** will generate a scatter plot like that shown below, where each point represents a pair of parameter values for each round of bootstrap or Monte Carlo simulation. The vertical and horizontal red lines mark the best-fit value for that parameter.



For this example (from a different fit) the data points are clearly clustered in an elliptically-shaped region that runs diagonally from upper left to lower right. This pattern indicates that the mass values for species 2 are moderately correlated with the fraction values for species 2, with a negative correlation in this case. As the mass value is increased or reduced from the best-fit value it is possible to partially compensate for that change, and minimize the increase in the *variance* of the fit, by simultaneously decreasing the fraction parameter. That is, the two parameters cannot be determined completely independently (although the correlation is not too high in this example).

If the points are very tightly clustered along a line (corresponding to a correlation coefficient near +1 or -1) this is an indication that too many parameters are being fitted (the information content of the data is not sufficient to determine all the parameters independently). In some cases the non-linearity of the fitting may produce clustering along a curve rather than a line. In any situation where the distribution is very tightly clustered along a diagonal line or a curve, the fitting results are probably not trustworthy and it may be appropriate to reduce the number of parameters by introducing **parameter constraints (Section 3.2.3.7)**, or by reducing the number of species.

If these parameters were not significantly correlated then the points would cluster along an ellipse that is oriented parallel to either the vertical or horizontal axis. (If the distribution was truly random the shape of the pattern would in principle be circular around the best-fit values if the axes were scaled in units of standard deviations, but in general the graph scaling will produce an ellipse rather than a circle).

This graph provides information that is quite similar to the cross-correlation coefficients that can be viewed using the **View parameter cross-correlations button (Section 3.2.6.3)**. However those values are based on a linear statistical model and assume the noise in the fitted data is random. The cross-correlation information obtained from these plots of bootstrap or Monte Carlo results is independent of such statistical assumptions.

Different pairs of parameters can be selected using the two drop-down list boxes. (Obviously when the same parameter is chosen for both axes the result will be a diagonal line.) The size of the data points on this graph can be altered by clicking on any of them with the mouse; the default size is automatically reduced as the number of data points increases to try to avoid too much overlap on the screen, but this choice may not be optimum when the graph is exported or printed.

Other features of this dialog box

The **Copy button** will place a copy of the current graph onto the Clipboard (in Windows enhanced *metafile* format).

The **Print button** will send this graph to the printer.

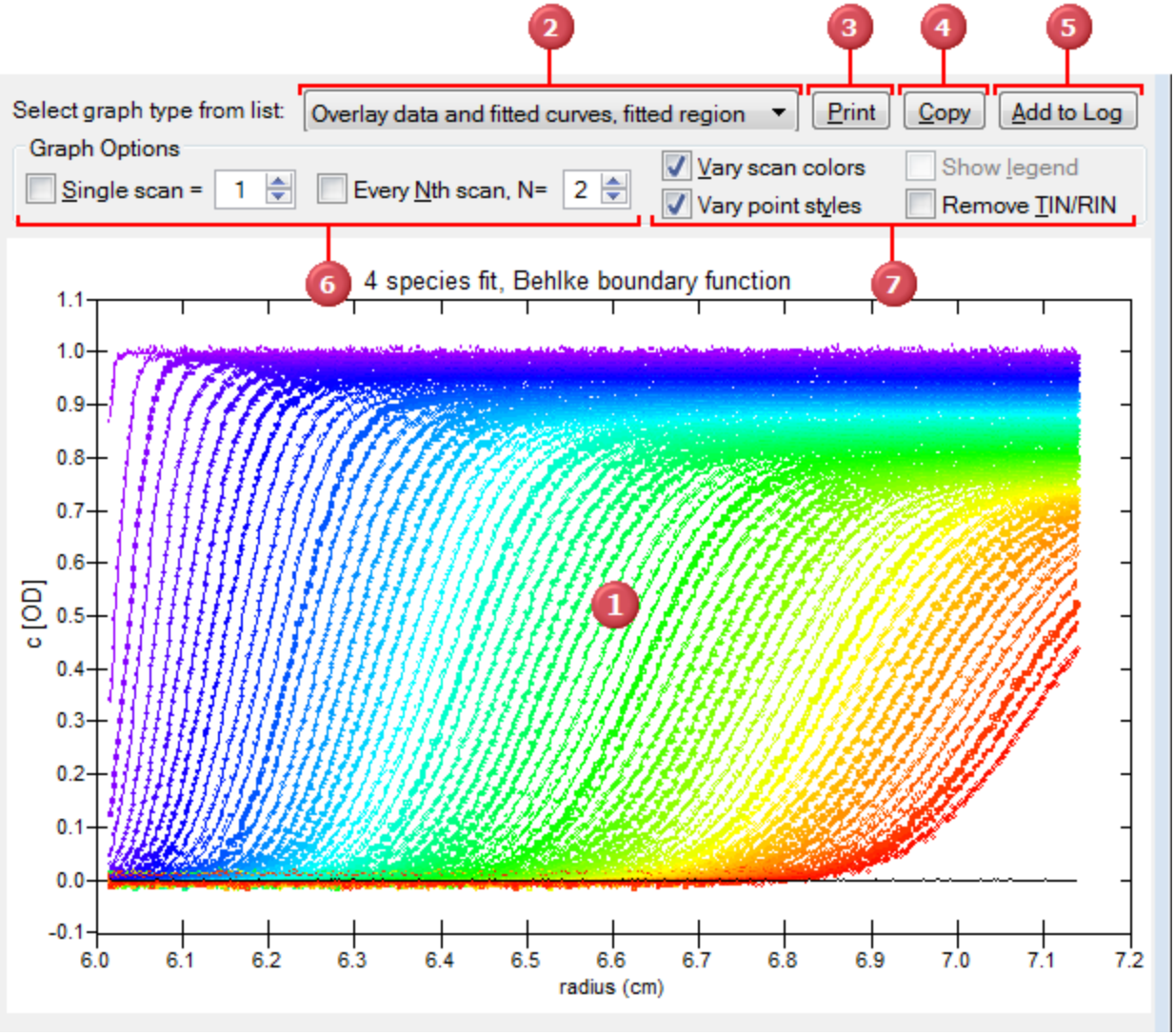
The **Add to log button** will add the current graph at the end of the **Log page (Section 3.1.3.2)**.

The **Readout data values checkbox** will turn on a readout of the graph coordinates corresponding to the current mouse position. This could for example be used to find the lower and upper 90% confidence limits for the actual distribution (without assuming a Gaussian shape) by locating the mass at the points where the cumulative frequency (magenta line) reaches 5% and 95%, respectively (giving the desired 10% probability that the true value lies outside the confidence interval).

3.2.7 Graph fit results page

This analysis page displays nine different types of graph of the fit results, fitted data, or theoretical fitted curves. The type of graph is selected from a drop-down list.

Each numbered area or button on the image below is described briefly below the image. The overlay graph shown is the default initial graph displayed for each new **analysis window (Section 3.1.3)**.



1 **Graph body** - This area holds the various types of graphs that are generated.

2 **Drop down list of graphs** - The different types of graphs are selected using this drop-down list box.

3 **Print button -** This button will bring up a standard Print dialog for selection of the printer to be used to print the current graph (full page). The graph orientation (portrait or landscape) and other printer-specific options can be changed using the Setup... button on the Print dialog.

4 **Copy button** - This button copies the current graph image to the Windows Clipboard as a high-resolution *metafile* so it can be pasted into another document.

5 **Add to log button** - This button will add the currently displayed graph at the end of the **Log page (Section 3.1.3.2)**.

6 **Scans shown options** - This region includes controls which govern which of the scans in the fit are shown on the graph:

- For graphs which normally show all the fitted scans (such as the overlay plot shown above), it is possible to show only a single specific scan by checking the **Single scan checkbox**. Which scan is shown (from first to

last among those being fitted) is then governed by the adjacent up/down spinner control.

- For graphs which normally show all the fitted scans (such as the overlay plot shown above), it is possible to show only every Nth scan (to reduce crowding and overlap on the graph) by checking the **Every Nth scan checkbox**. The value of N is then controlled by the adjacent up/down spinner control.

7 **Other graph options** - This region includes controls which may change or disappear depending on which type of graph is displayed. These controls govern options for features such as the use of color, whether the point style is the same for all scans or varied, whether a graph legend is displayed, and for removing the fitted TIN and/or RIN from the raw data (to show clean boundaries).

Details about these options

The options which may be available are:

Vary scan colors or **Use color**

This check box essentially governs whether the graph is displayed in color or black-and-white.

Vary point styles

When checked the raw data points for different scans will be displayed using a sequence of 12 different symbols (to help distinguish one scan from another), starting with a circle. If un-checked, all scans will be shown as a '+' symbol.

When the number of data points per scan is large using different symbols causes points to overlap and a cluttered graph, so by default when the points per scan exceeds 1000 (as is normally the case for interference scans) only a '+' is used for all scans and this option is disabled.

When the number of scans is large using just a '+' may give a graph that is cleaner and with less overlap between scans.

Show legend

This check box governs whether a legend is displayed below the graph to provide a description of each different data set in the graph. For scan data the legend entry is the file name of that scan.

Legend display can also be toggled on or off by right-clicking on the graph to bring up the graphing pop-up context menu and using the 'Hide legend' entry within the 'Legend style' sub-menu.

By default for each new graph the legend is suppressed when the total number of legend entries exceeds 20 because the legend takes up considerable space and gets harder to read when the number of entries is high. This behavior can be over-ridden using this **Show legend checkbox**.

The third-party graphing package used by SVEDBERG will only display up to ~64 total entries in the legend (the actual number depends on the content also) and thus by default this check box is disabled when the total entries exceeds 64. However under some circumstances a legend with 64 or less entries may not actually appear even when the graphing package is told to display it.

Remove TIN/RIN

When the data analysis model includes **fitting of the TIN and/or RIN (Section 3.2.3.5)** checking this box will subtract this systematic noise from the data being displayed in the graph, which can be useful to get a cleaner display (one that is often easier to interpret).

Spread residuals

This option is displayed only for the line graph form of residual plots and replaces the 'Remove TIN/RIN' option. Selecting this option will cause a constant vertical separation between the residuals from each successive scan, which reduces the overlap and makes it easier to see details for individual scans.

Unfortunately this type of display only works well when the total number of scans being displayed is roughly 50 or less. Therefore by default this option is turned off for each new fit with more than 50 fitted scans. For scans with more than 50 fits it may be useful to first reduce the number of scans being displayed with the **Every Nth scan checkbox** and then turn on **Spread residuals**.

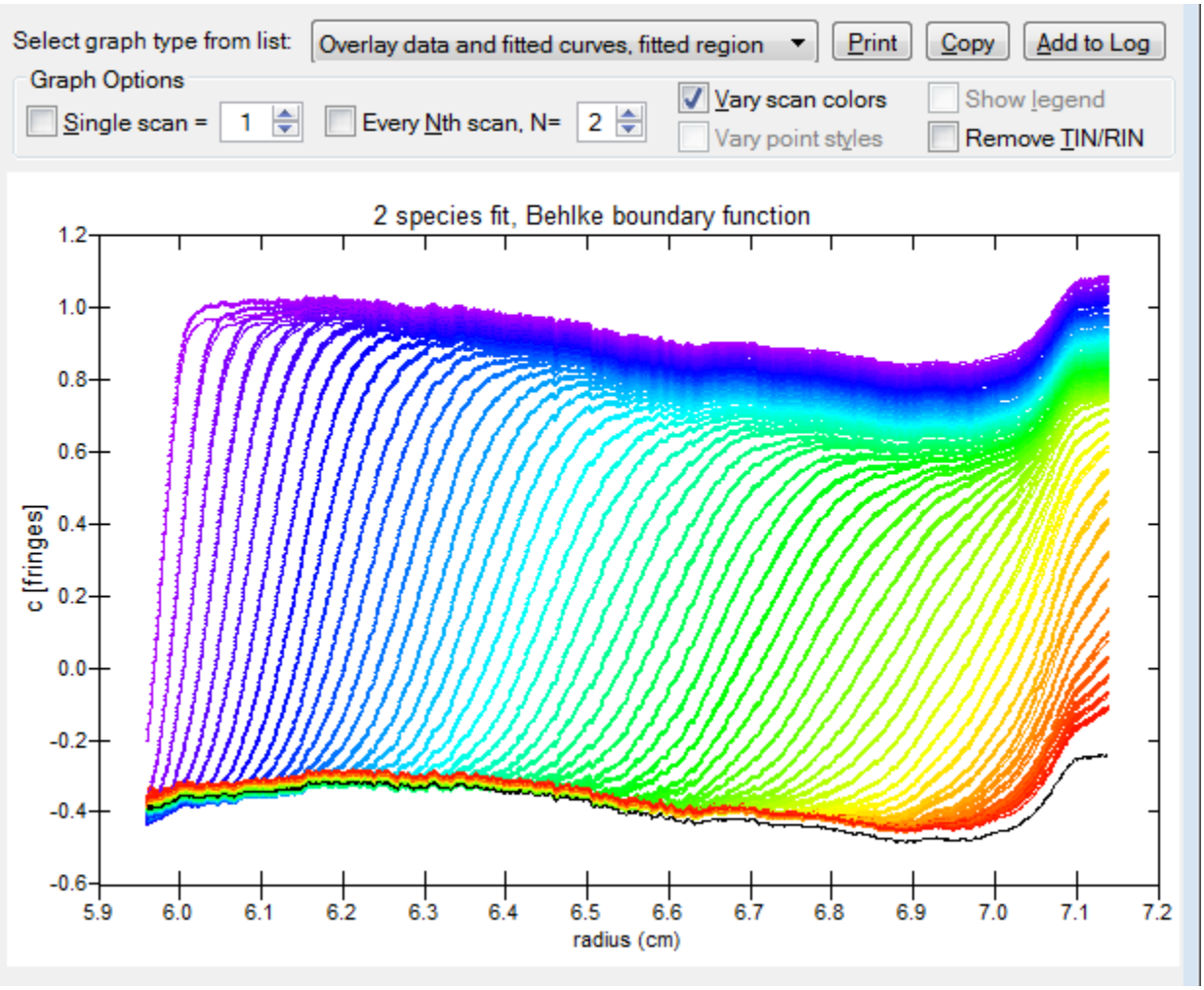
Graphs available

The drop-down sections below show examples of each type of graph (not necessarily from the same analysis), with brief comments about the intended use of that plot type. Note that although these descriptions refer to the fitted data, if you have reached this page after doing a **simulation (Section 3.2.5.3)** rather than a fit then the plots show the results of the simulation.

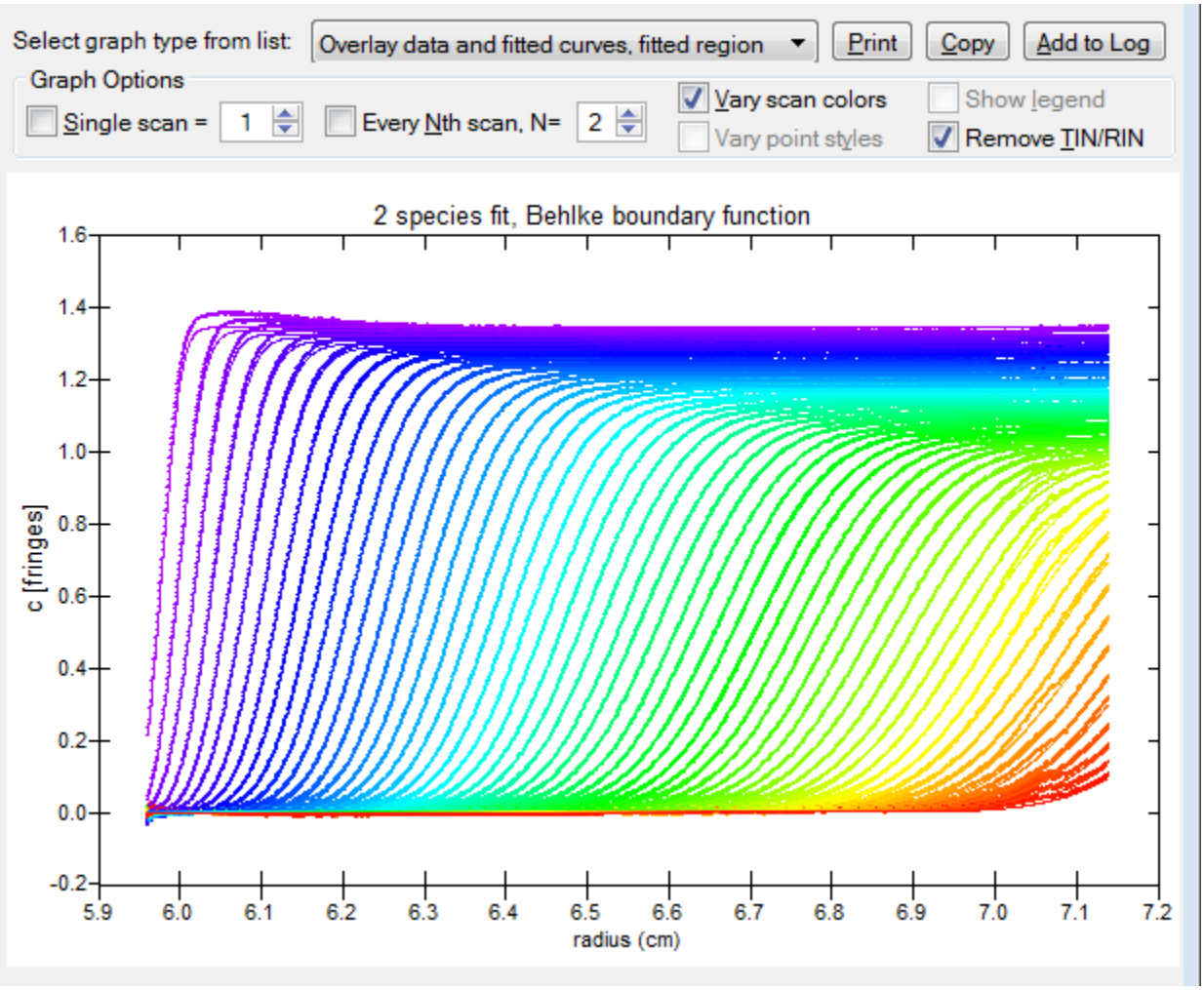
1. Overlay data and fitted curves, fitted region

Below is an example of this type of plot from a fit of some interference scans. It overlays all of the data being fitted (shown as individual data points) with the theoretical curves from the fit (shown as continuous lines). Since the TIN was included in the fit then by default it will be displayed as a black line.

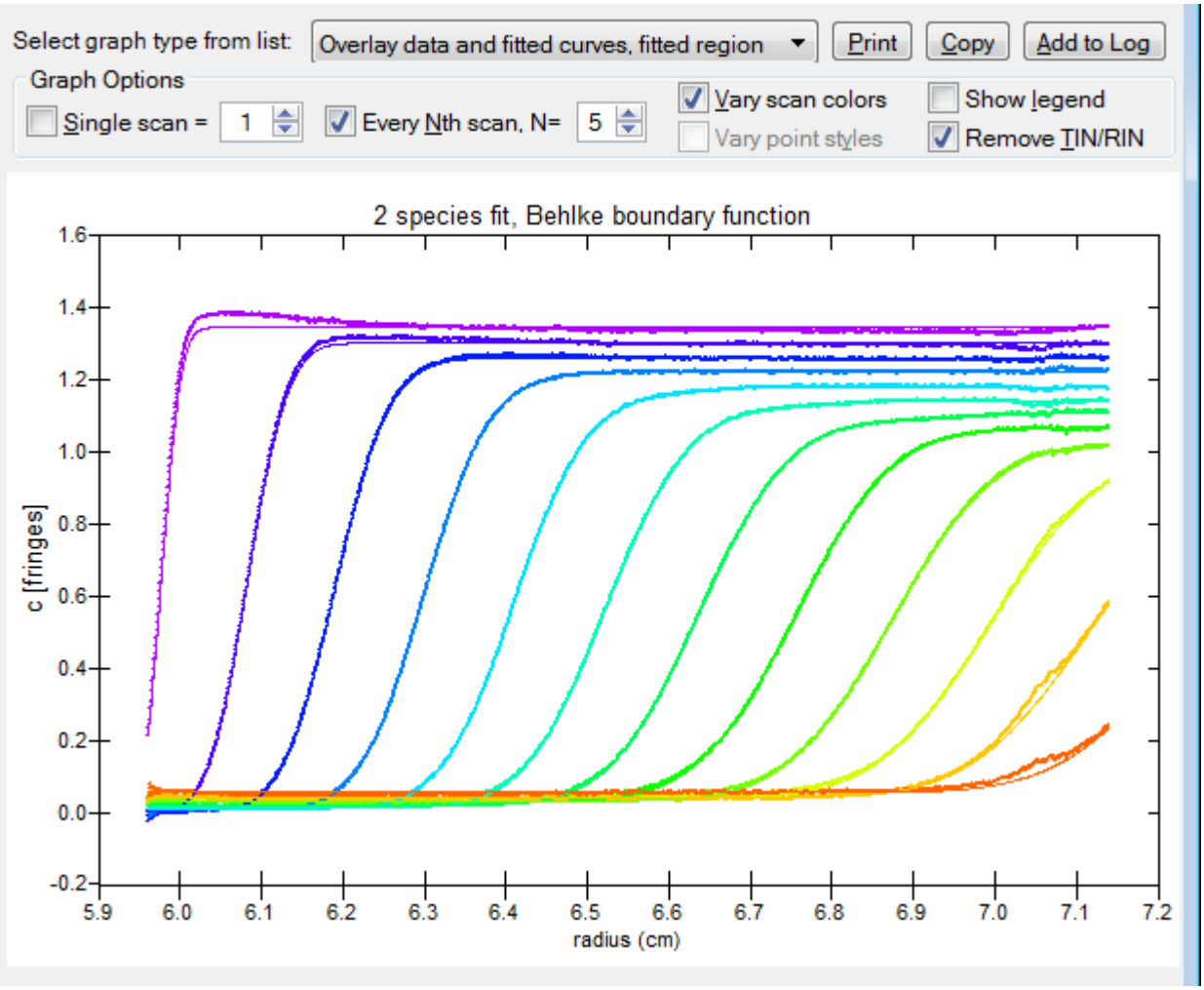
This plot summarizes the data being fitted and whether the fit is a good representation of that data. This example is not a particularly good fit and clearly the data near the meniscus deviate from the fit.



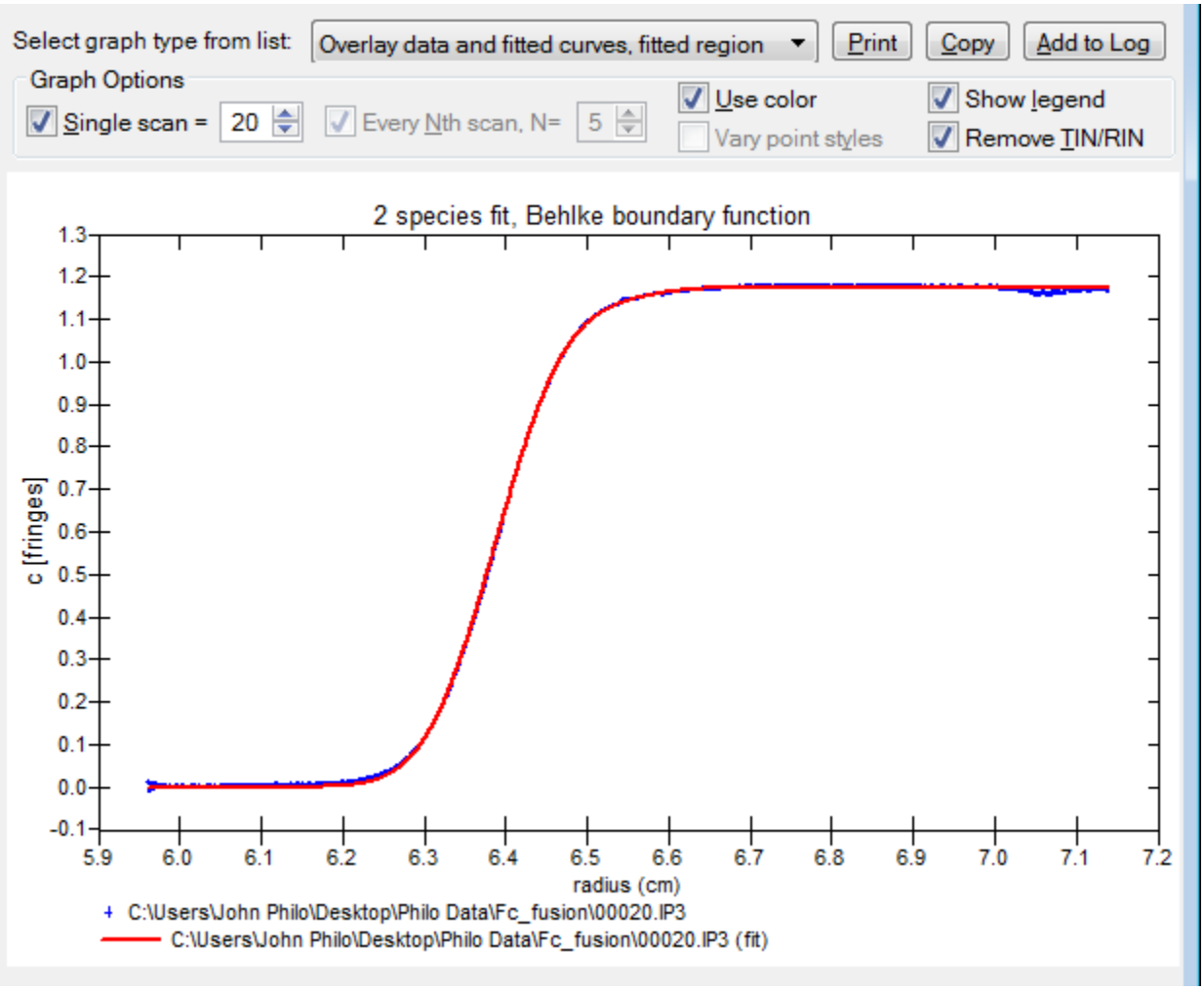
It is also sometimes useful to remove the computed TIN and/or RIN so you can see the data and fits without the systematic noise, which can be done by checking the **Remove TIN/RIN** checkbox, as shown below.



When there are many scans it can be hard to see whether the data and fit agree. Thus it is also sometimes helpful to show only a portion of the scans (which also may be better for presentations or publication). The image below shows the result when only every fifth scan is shown by using the **Every Nth scan** option.



An overlay with only a single scan can be obtained by checking the **Single scan** checkbox as shown below, and you can then scroll through all the scans using the up/down spinner.



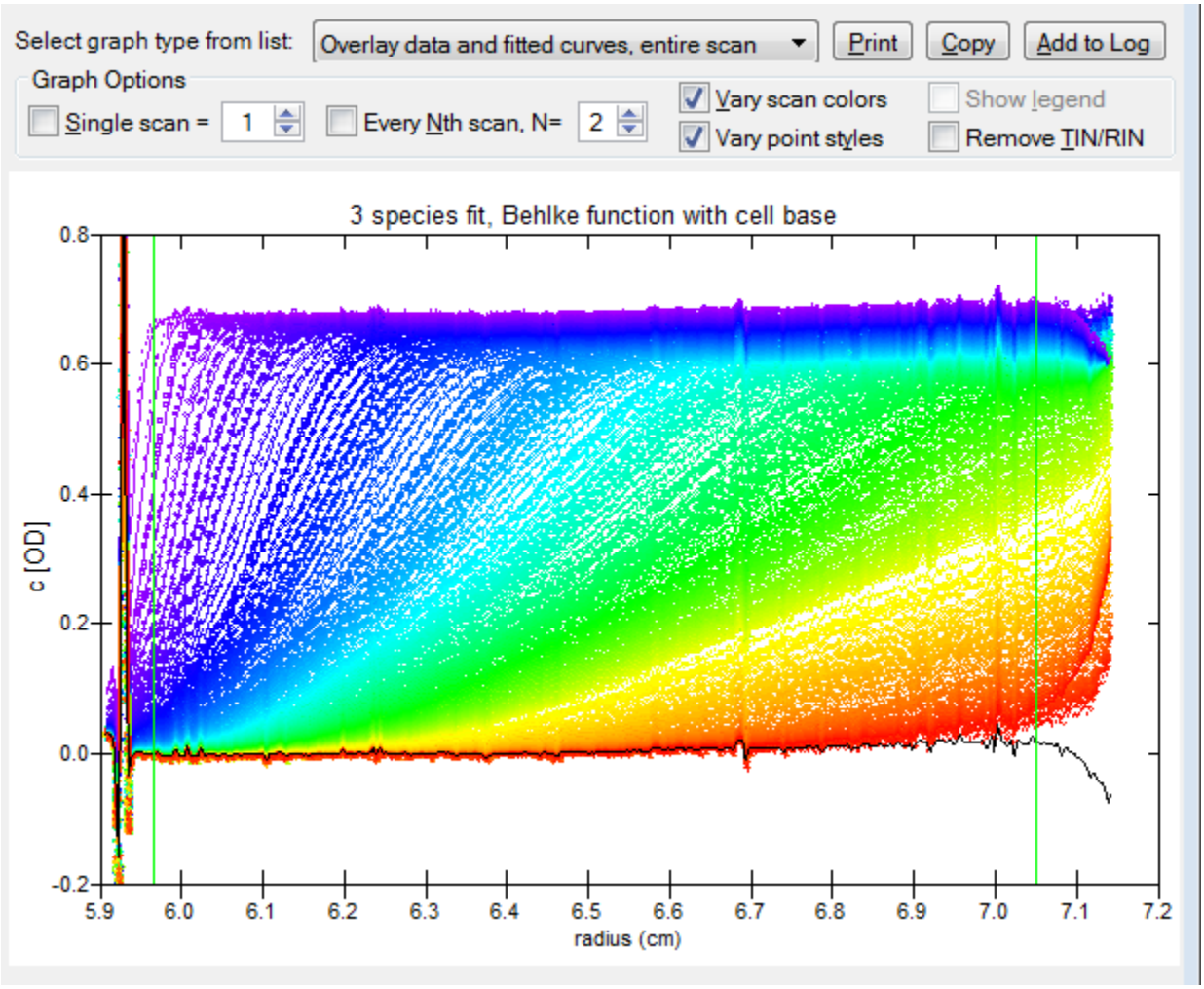
For more information about other graph options controls available with this plot see the **Details about these options** drop-down section under point 7 above.

2. Overlay data and fitted curves, entire scan

The differences between this plot and #1 are:

1. it displays all the entire scans (not only the portions actually being fitted)
2. the best-fit theoretical curves are also calculated and displayed over the entire raw data range
3. if the TIN has been fitted it is also calculated over the entire raw data range
4. the end points of the fitting region are marked by green vertical lines

An example is shown below.



One important advantage of this type of plot is illustrated above. Note that the TIN (black curve) starts to curve downward significantly just past the end of the fitting region at 7.05 cm (marked by the green vertical line). Since this type of curvature is not something we expect for the true baseline of an absorbance scan, this feature suggests a problem with the fitting model: the model is not correctly reproducing what is happening near the base of the cell. In this case the downward curvature indicates that the model is producing too much buildup at the cell base. That could indicate that the position of the cell base needs to move outward, but a similar situation might arise if the macromolecule is precipitating at the cell base.

The fact that the deviation is clearly seen just beyond the fitting limit indicates that these problems probably are influencing some of the data points being fitted also. Therefore it would be wise to either move the end of the fitting region to the left, and/or move the cell base position outward (or possibly to remove the buildup at the cell base entirely by changing the fitting function to the **Behlke boundary function (C:\Users\Public\Documents\Visual Studio Projects\Svedberg Help Source\{4141547C-6F91-46E6-87B2-403D7525DA21})**).

As discussed in **Should I fit the cell base position? (C:\Users\Public\Documents\Visual Studio Projects\Svedberg Help Source\{BEE314E3-0C09-4B1D-A93B-AD0C0C4FC939})**, it is seldom possible to correctly model a substantial buildup near the cell base, and therefore in general it is a good idea to avoid having your fit substantially influenced by the details of what is happening there. For absorbance data we expect the TIN to be fairly flat, so this type of plot can serve as a 'reality check' for whether there may be problems near the cell base. If the TIN seems flat for a region beyond the end of the

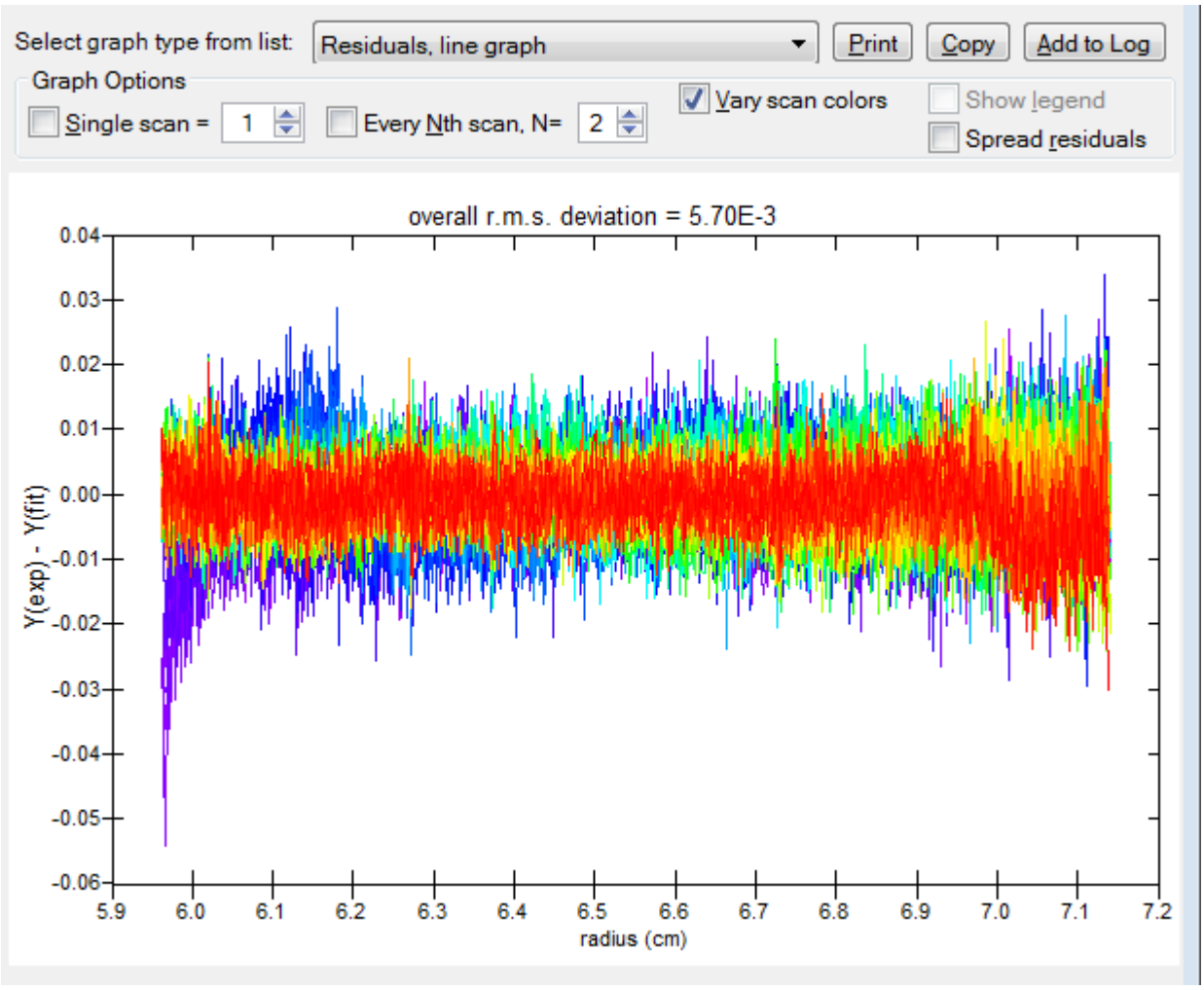
fitting region that is a good sign that the base of the cell is not causing significant problems. Unfortunately this test is not so easily applied for interference scans.

This type of plot may be more useful than #1 for presentations or publications because it shows all the raw data. Like #1, when the TIN and/or RIN have been fitted they can be removed to clean up the graph by checking the **Remove TIN/RIN** checkbox (see examples for graph #1).

If you wish to remove the green vertical lines from the graph that can be done by right-clicking on the graph with the mouse and then un-checking the 'Show Annotations' item in the pop-up context menu.

3. Residuals, line graph

This graph shows the residuals for each fitted scan as a line graph. In the example below the scans are all overlaid, and when the number of scans is large it may be difficult to observe small systematic deviations. Nonetheless in this case the 'rainbow colors' help to show that there are systematic deviations near the meniscus for the early scans (violet), and systematic deviations near 7-7.1 cm for the last scans (red)



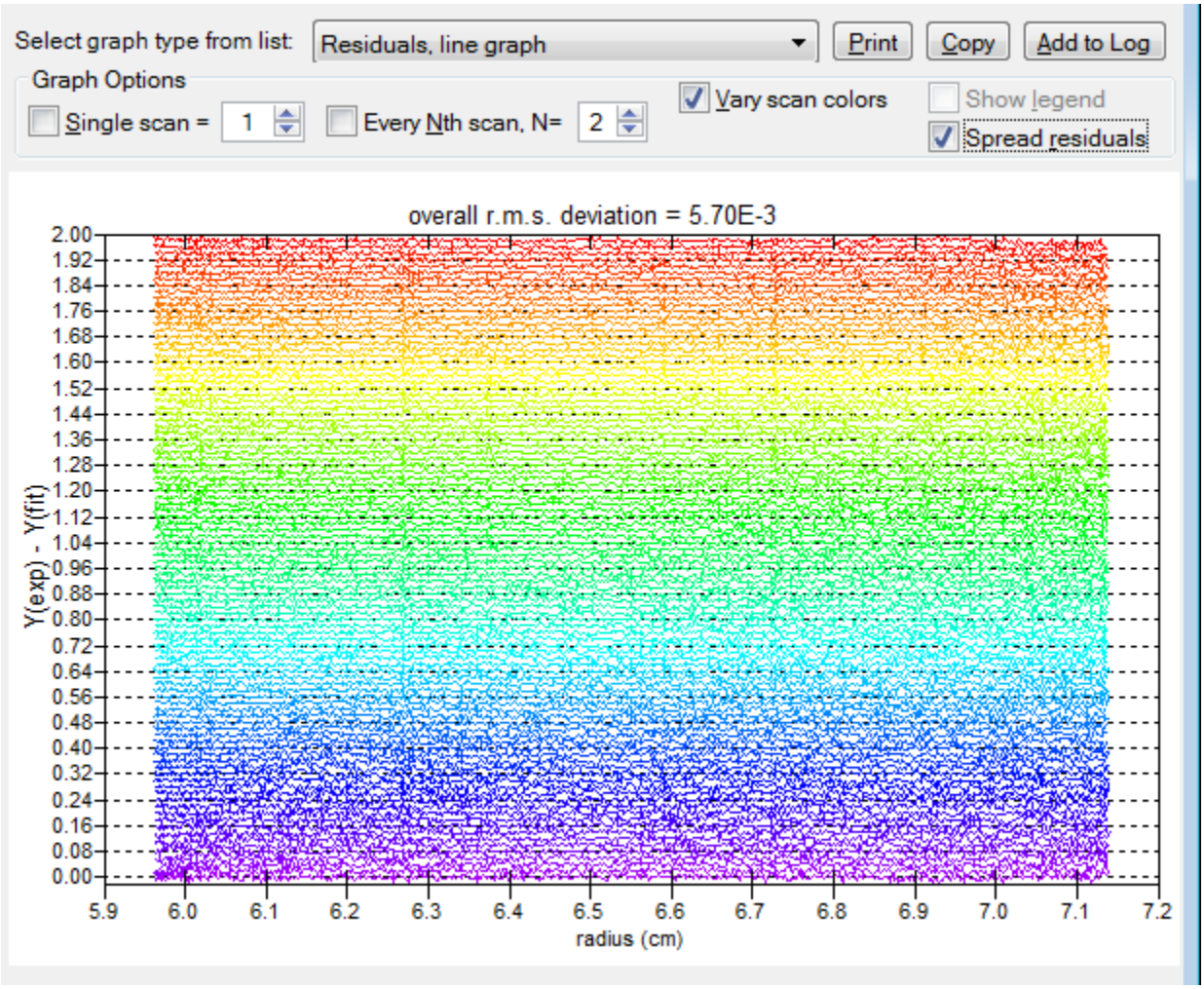
This residual plot is functionally equivalent to, but formatted differently than, the one that is shown in the lower panel of the plot shown on the Graphical fit monitor tab of the **Perform least-squares fit page (C:\Users\Public\Documents\Visual Studio Projects\Svedberg Help Source\hs1240.html)**.

The value of the *rms residual* for this fit is displayed as the plot title.

Note that if you prefer alternate plotting methods, such as plots with points connected by lines, or stick plots, the plotting method can be quickly changed by right-clicking on the graph and using the

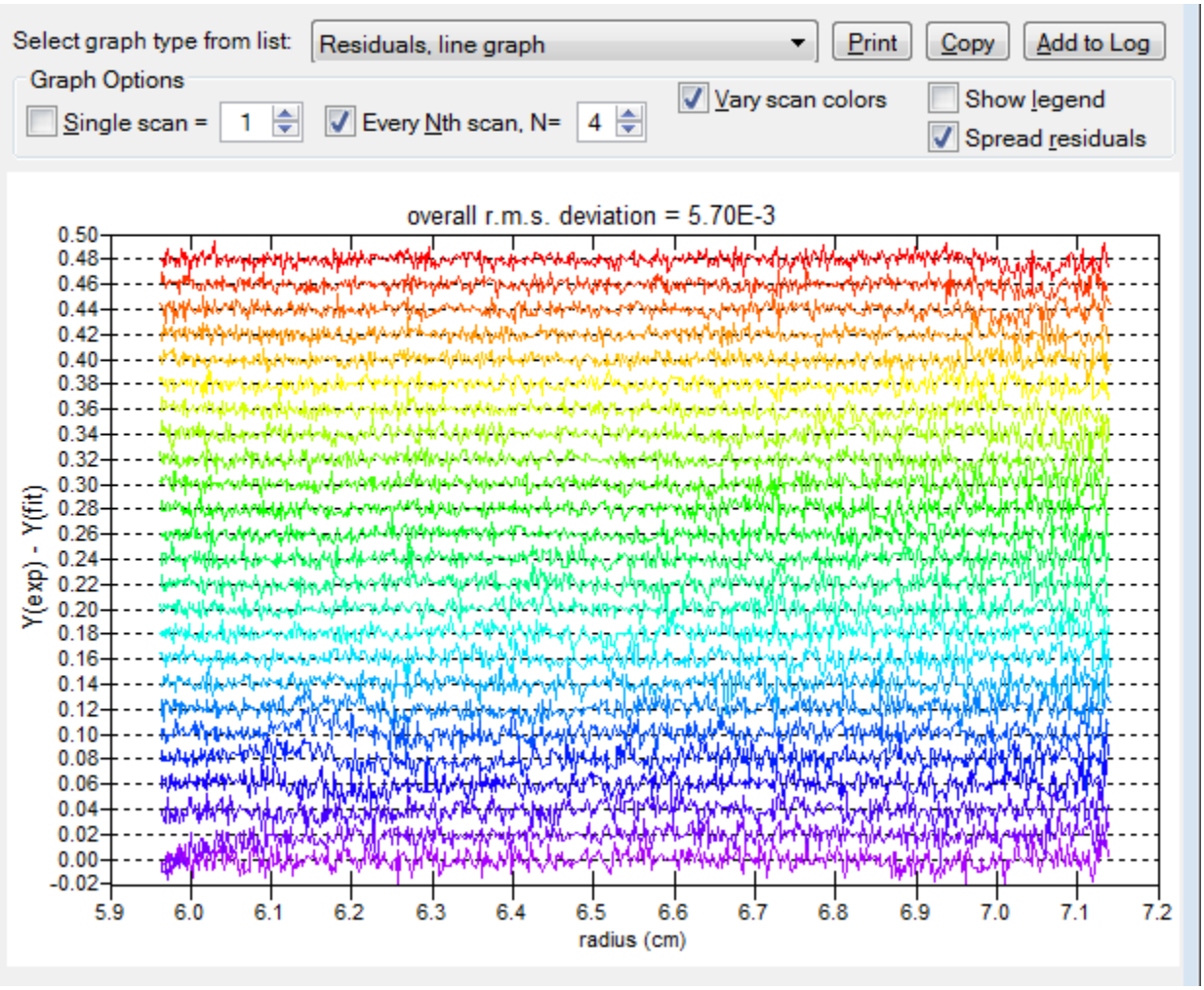
Plotting Method sub-menu on the pop-up graph context menu.

The **Spread residuals** option is shown only for this graph. This option will reduce the overlap between the residuals from different scans by separating each successive scan by a constant vertical offset. However this option usually only works well when the number of fitted scans is roughly 50 or less, as can be seen in this example below from the fit shown above of 100 absorbance scans.



Note that for the case above only the vertical positions for every third scan are marked by grids and tic labels.

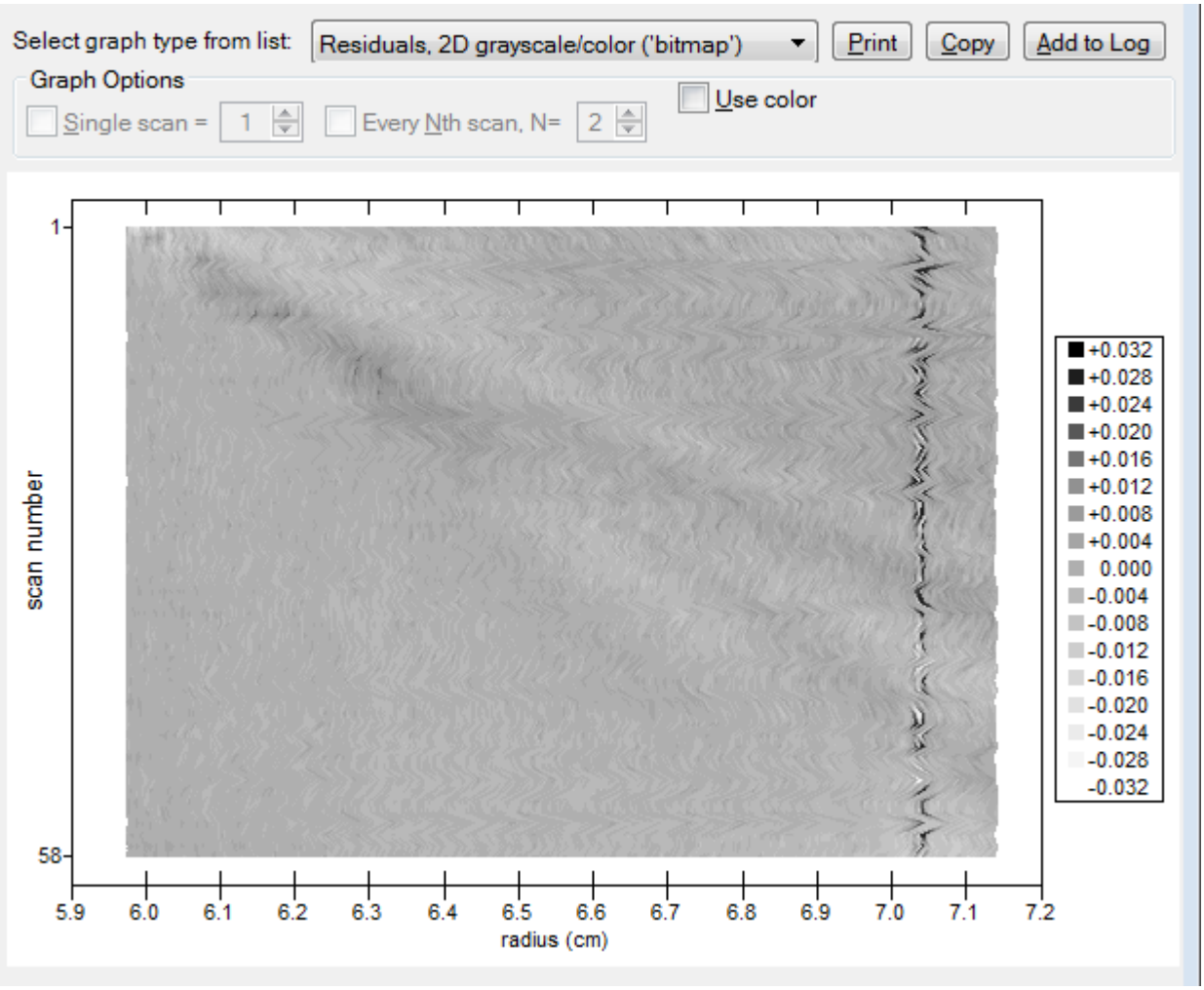
With this fit when the number of residual plots displayed is reduced to 25 by using the **Every Nth scan** option, as shown below, the black horizontal grid lines provide a zero reference for each scan and it becomes easier to see that systematic residuals are present between ~6.0-6.3 cm for about the first third of the run (violet and blue colors) and another systematic pattern shows around 7.0-7.1 cm late in the run (red).



4. Residuals, 2D gray scale/color 'bitmap'

In this representation of a residual plot the residuals for all the scans are shown in a 2-dimensional contour plot with radius along the X axis and scan numbers along the Y axis (with the first scan at the top), and the magnitude of each residual determines the shade of gray or color shown at that position.

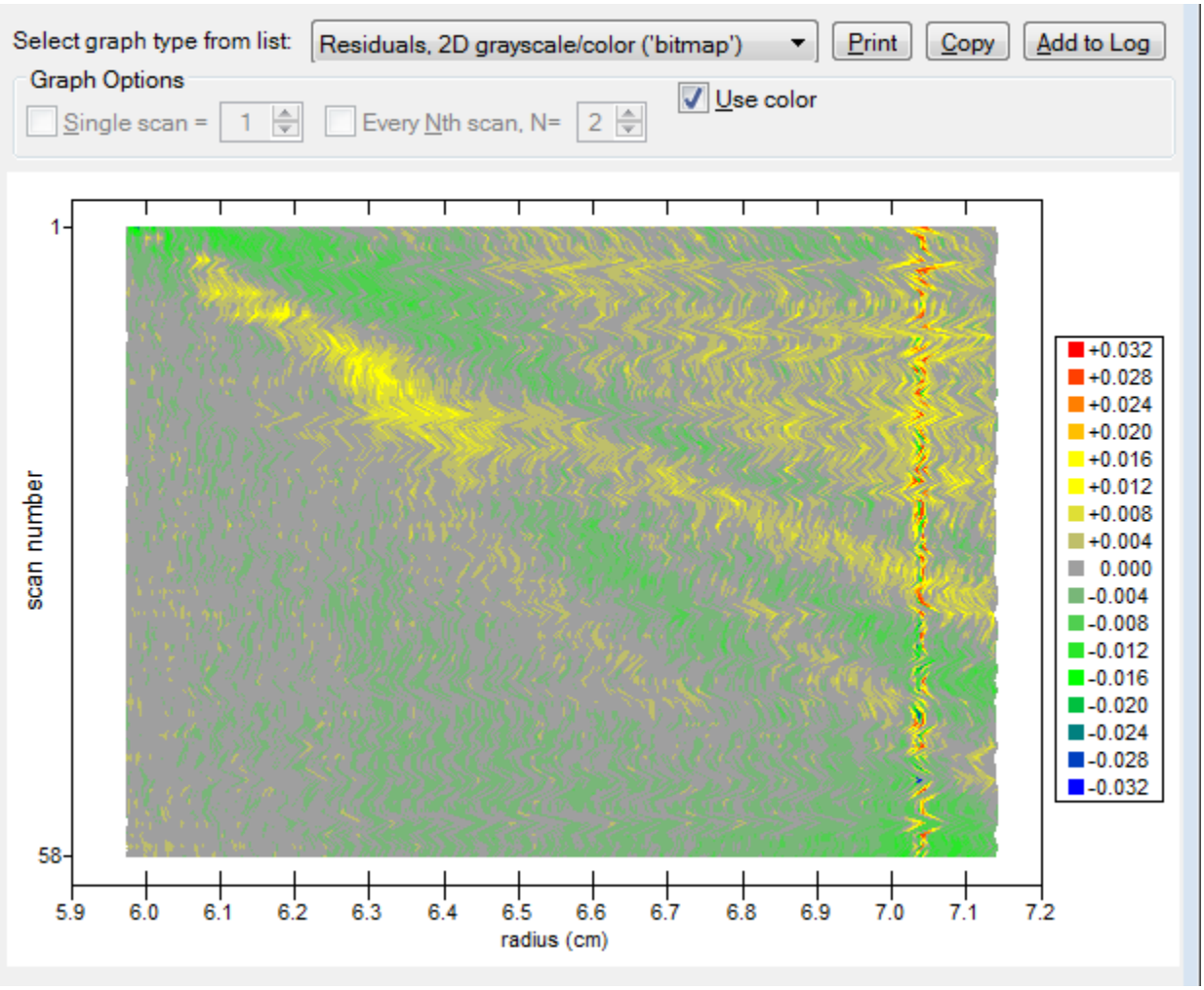
The principal advantage of this type of plot is that the human brain is very good at discerning subtle patterns among what is mostly random noise, and therefore these plots are quite sensitive to whether there are any systematic patterns in the residuals. If the residuals are truly random then the plot should be fairly uniformly gray. In the example below in the upper portion of the graph (earlier in the run) there is a darker feature running diagonally downward, moving along with the main boundary, with a parallel faint lighter region above it. Later in the run (bottom half of the graph) these features split to form a fork-like pattern of lighter and darker diagonal zones. There is also an obvious vertical feature near 7.04 cm due to something on the cell window at that position (in other words, the fitted TIN was not completely successful at removing the systematic noise).



This plot is similar to the bitmap plots in SEDFIT, and serves the same purpose, but differs in several ways:

1. In the SEDFIT bitmap plots, each scan corresponds to exactly one vertical pixel on the computer screen. That means the vertical size of the graph is determined by the number of scans being fitted. The vertical size of the SVEDBERG plots is a constant 94% of the vertical axis extent, and the scan numbers of the first and last scans are marked on the Y axis. The aspect ratio of the graphs can be changed by simply resizing the analysis window.
2. The SEDFIT plots are true bitmaps at screen resolution and cannot be resized without introducing distortions. For absorbance scans the bit-wise mapping of radius values to individual bits also sometimes leaves gaps in the image (because the spacing between points varies), which show as black pixels. In contrast these SVEDBERG plots are true high-resolution 2D plots with interpolation, and can be copied to other programs and rescaled without loss of image quality. The drawback is that simple bitmaps can be generated much more quickly, so often there will be a delay of several seconds before the graph is displayed.
3. These SVEDBERG plots always have a key on the right showing the intensity scale, and the scaling can be changed by simply clicking within the key.

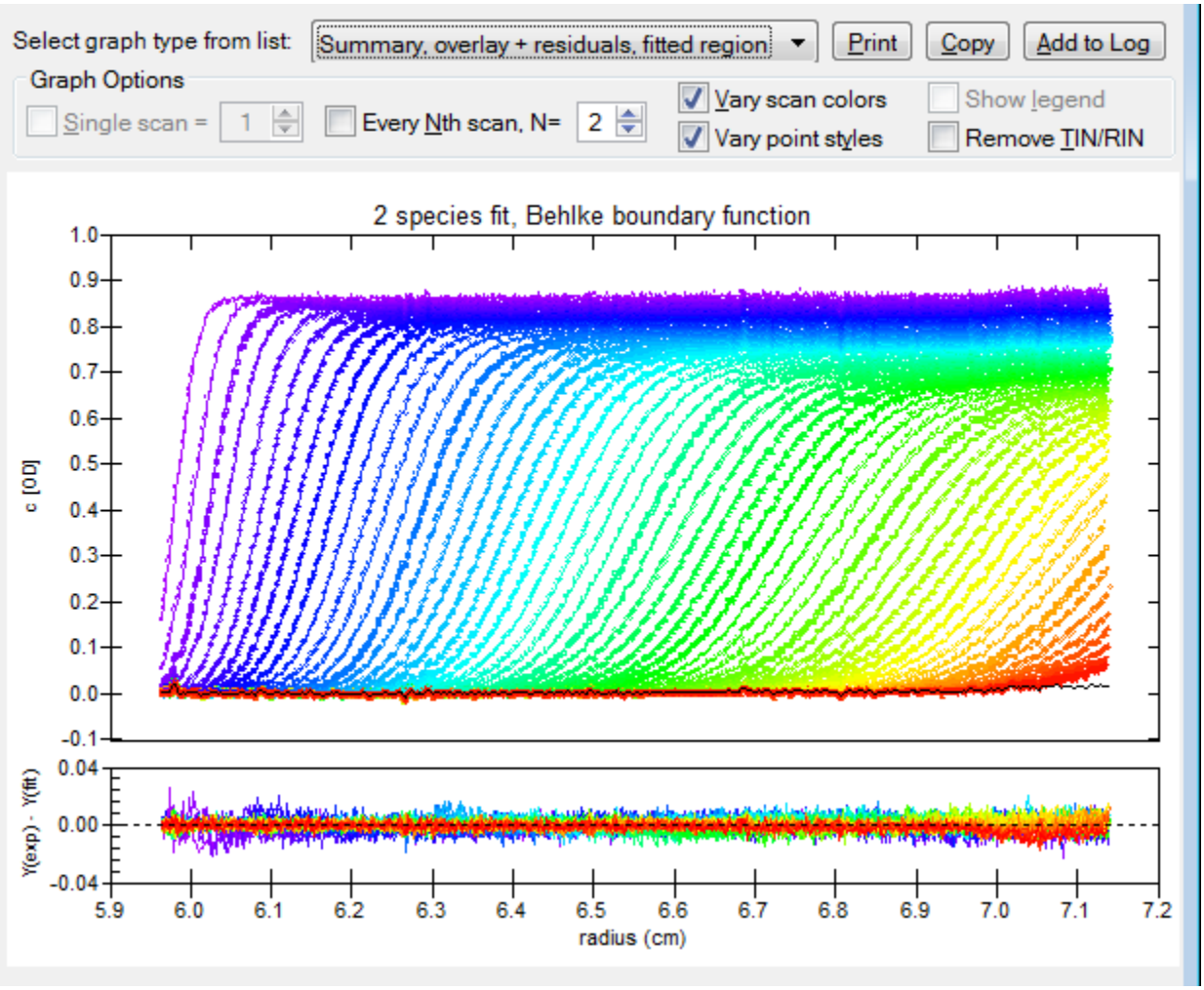
After each new fit the default is to display this plot in grayscale, but a color version with hot vs. cool colors to mark the positive and negative residuals can be obtained by clicking the Use colors checkbox. The same data shown above are shown in color below.



The default full scale range of +/- 0.032 for the residuals for absorbance or interference scans usually works well, but may need to be increased for noisy data or when the systematic deviations are large, or possibly decreased if the noise level is quite low. This can be done by clicking anywhere within the key on the right, which will bring up a small dialog box for entering the full scale value. Entering a zero will return to the default scale. Full scale values which are divisible by 8 such as 0.040 or 0.048 will result in round numbers for each color or shade of gray. For fluorescence scans the default full scale value is 8 times the rms residual.

- **5. Summary, overlay + residuals, fitted region**

This plot creates a summary of the fit by combining an overlay plot (#1 above) with a residual plot (#3) into a single graph object with two sets of Y axes. This plot is the same as the plot shown on the Graphical fit monitor tab of the **Perform least-squares fit page (Section 3.2.5)** and the one that is used in printed reports and automatically added to the **Log page (Section 3.1.3.2)**.



Due to limitations of the 3rd party graphing library used by SVEDBERG it is not possible to display this type of graph for more than 80 scans. Therefore if the fit uses more than 80 scans then when this graph is selected the **Every Nth scan** option is automatically turned on (but it is disabled since it must be on), and a minimum value of N is set which produces 80 or less displayed scans.

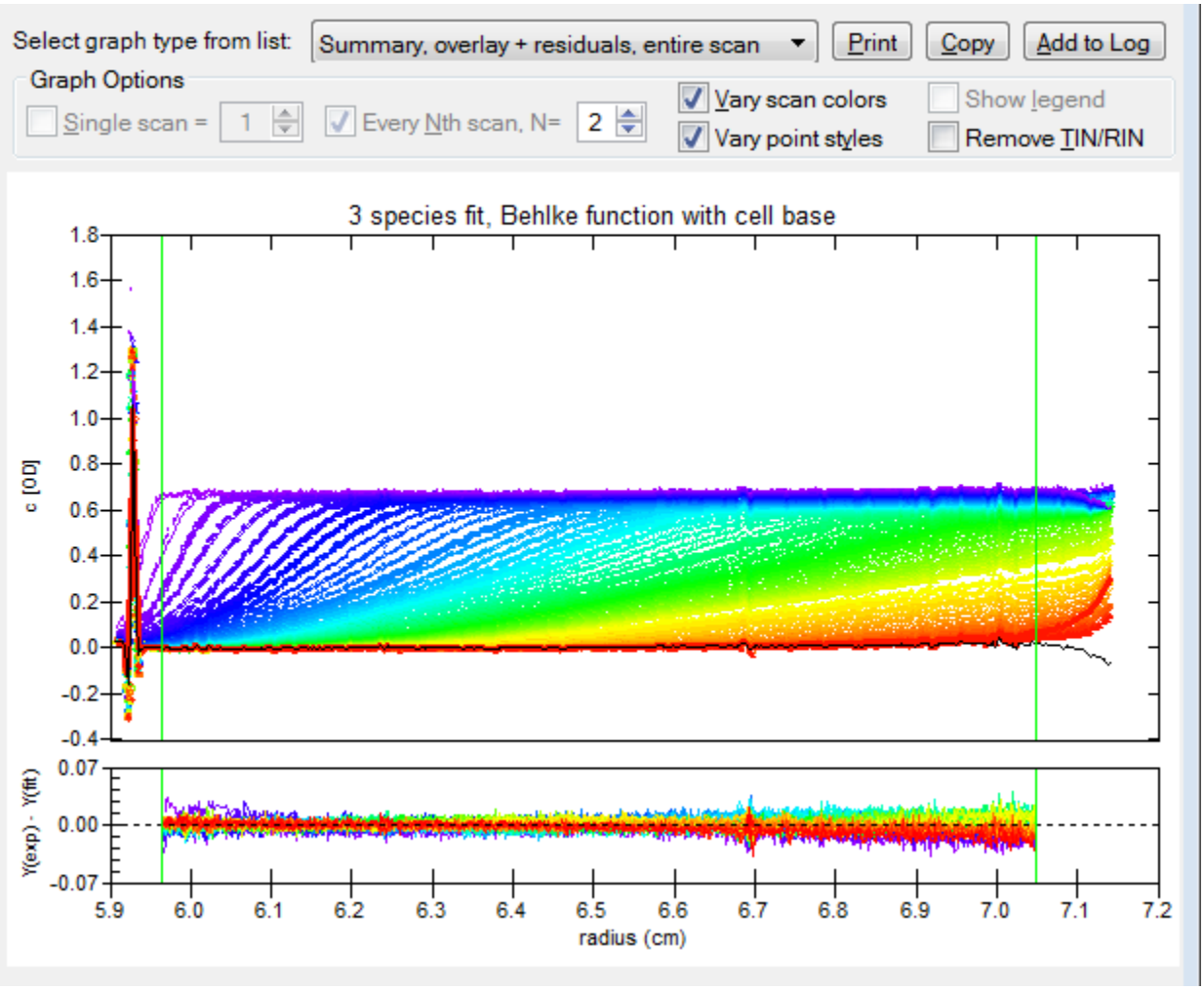
The relative vertical size of the overlay and residual plot portions can be altered using the mouse. When the mouse hovers over the region in between these plots the cursor changes to a vertical arrow and the division point between the graphs can be moved by clicking and dragging up or down with the mouse.

The options available for this graph are the same as those for #1 above except that the single scan plots are not available.

6. Summary, overlay + residuals, entire scan

This plot creates a summary of the fit by combining an overlay plot (#2 above) with a residual plot (#3) into a single graph object with two sets of Y axes. Unlike plot #5 this plot shows the full radial range covered by the raw scans, not only those points actually being fitted. The best-fit theoretical curves are also calculated and displayed over the entire raw data range, and if the TIN has been fitted it is also calculated over the entire raw data range.

An example is shown below.



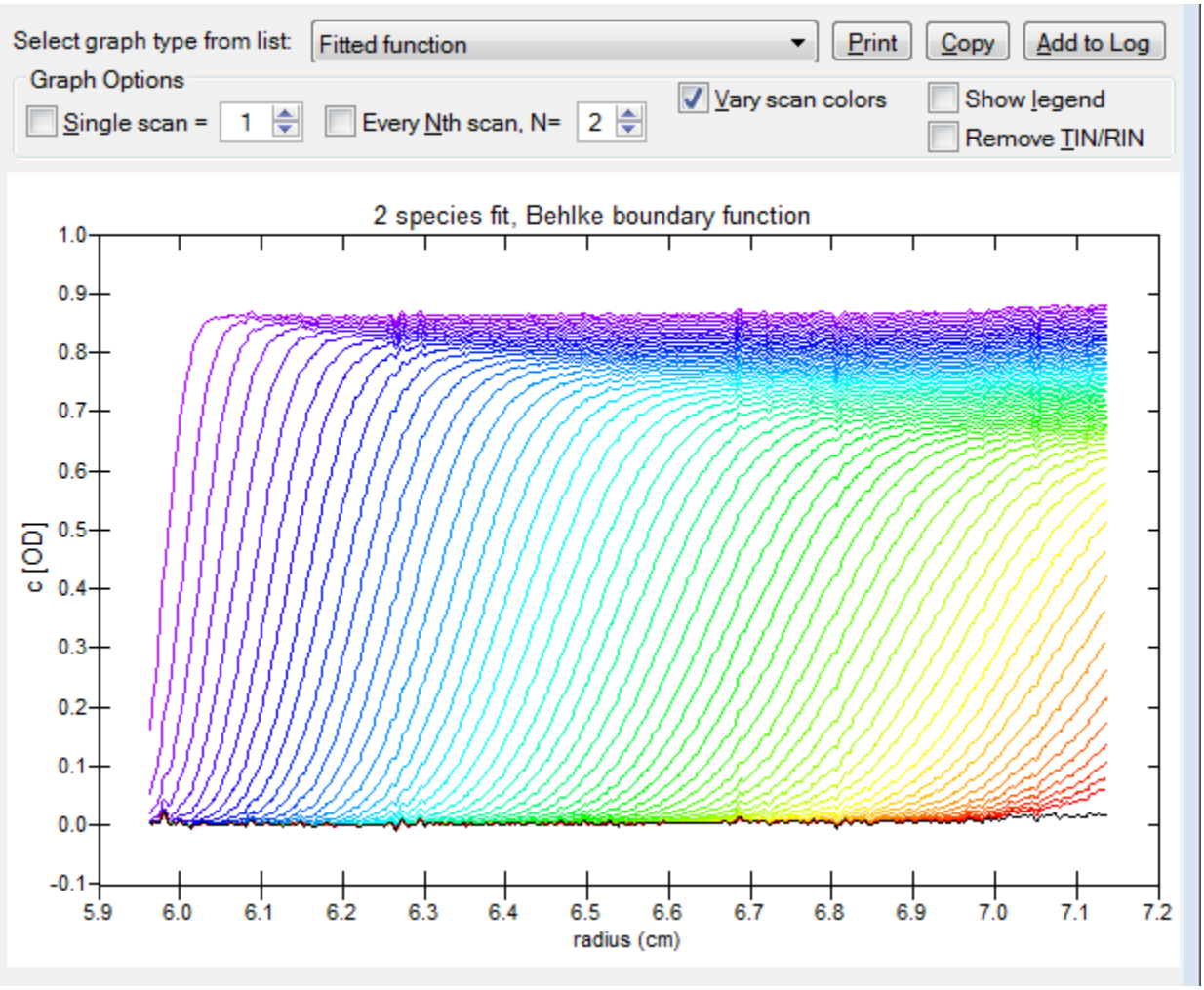
If you wish to remove the green vertical lines that mark the actual fitting region from the graph that can be done by right-clicking on the graph with the mouse and then un-checking the 'Show Annotations' item in the pop-up context menu.

Due to limitations of the 3rd party graphing library used by SVEDBERG it is not possible to display this type of graph for more than 80 scans. Therefore if the fit uses more than 80 scans then when this graph is selected the **Every Nth scan** option is automatically turned on (but it is disabled since it must be on), and a minimum value of N is set which produces 80 or less displayed scans.

The relative vertical size of the overlay and residual plot portions can be altered using the mouse. When the mouse hovers over the region in between these plots the cursor changes to a vertical arrow and the division point between the graphs can be moved by clicking and dragging up or down with the mouse.

The options available for this graph are the same as those for #1 above except that the single scan plots are not available.

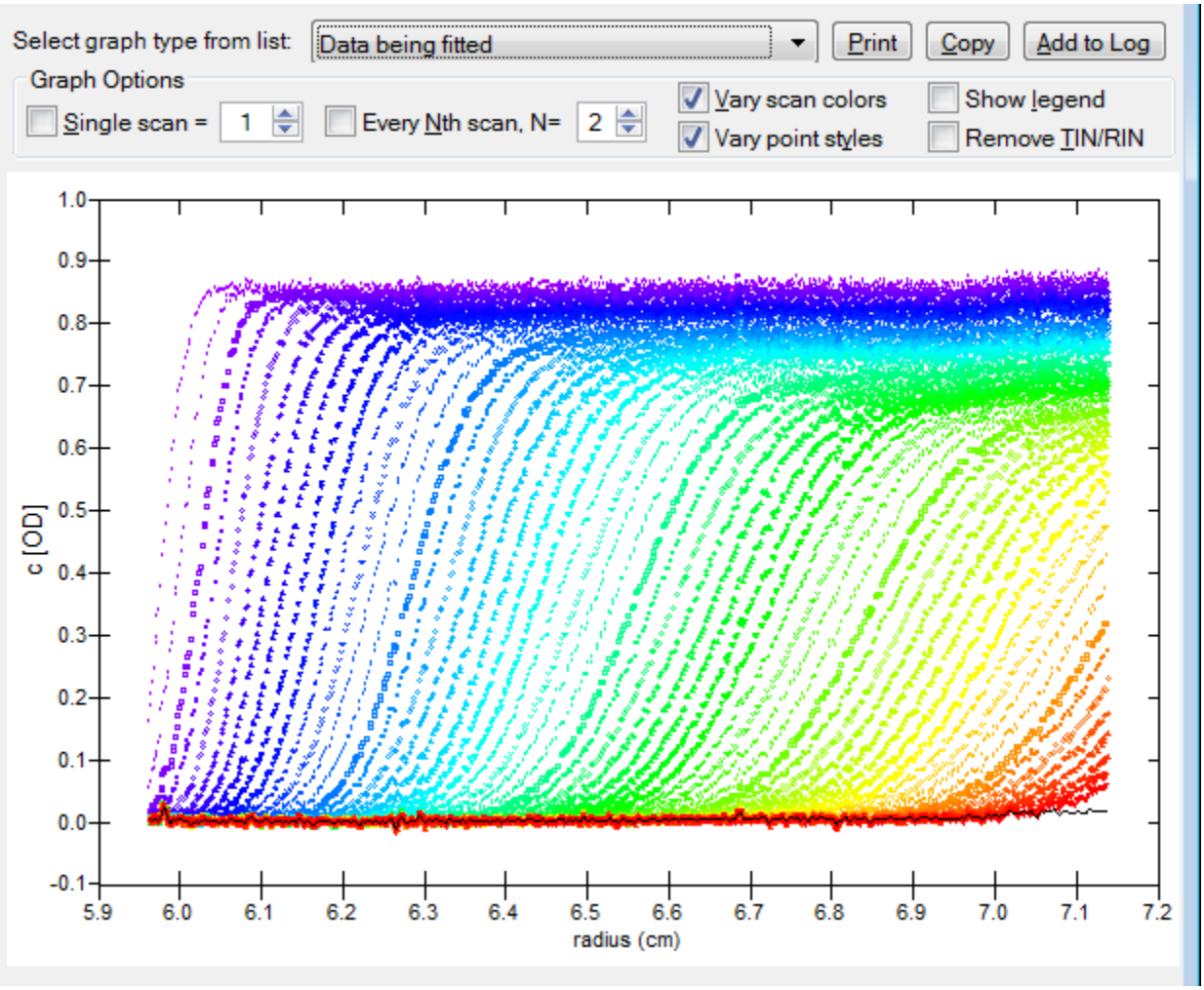
- **7. Fitted function**



This plot simply shows the best-fit curves for the data region that was actually fitted. When the TIN has been fitted, as in this example, it is added to the theoretical fitted function and also displayed as the black trace, but it can optionally be removed by checking the **Remove TIN/RIN** option.

Options are available for showing only a single one of the fitted scans, or for showing only every Nth scan.

- **8. Data being fitted**



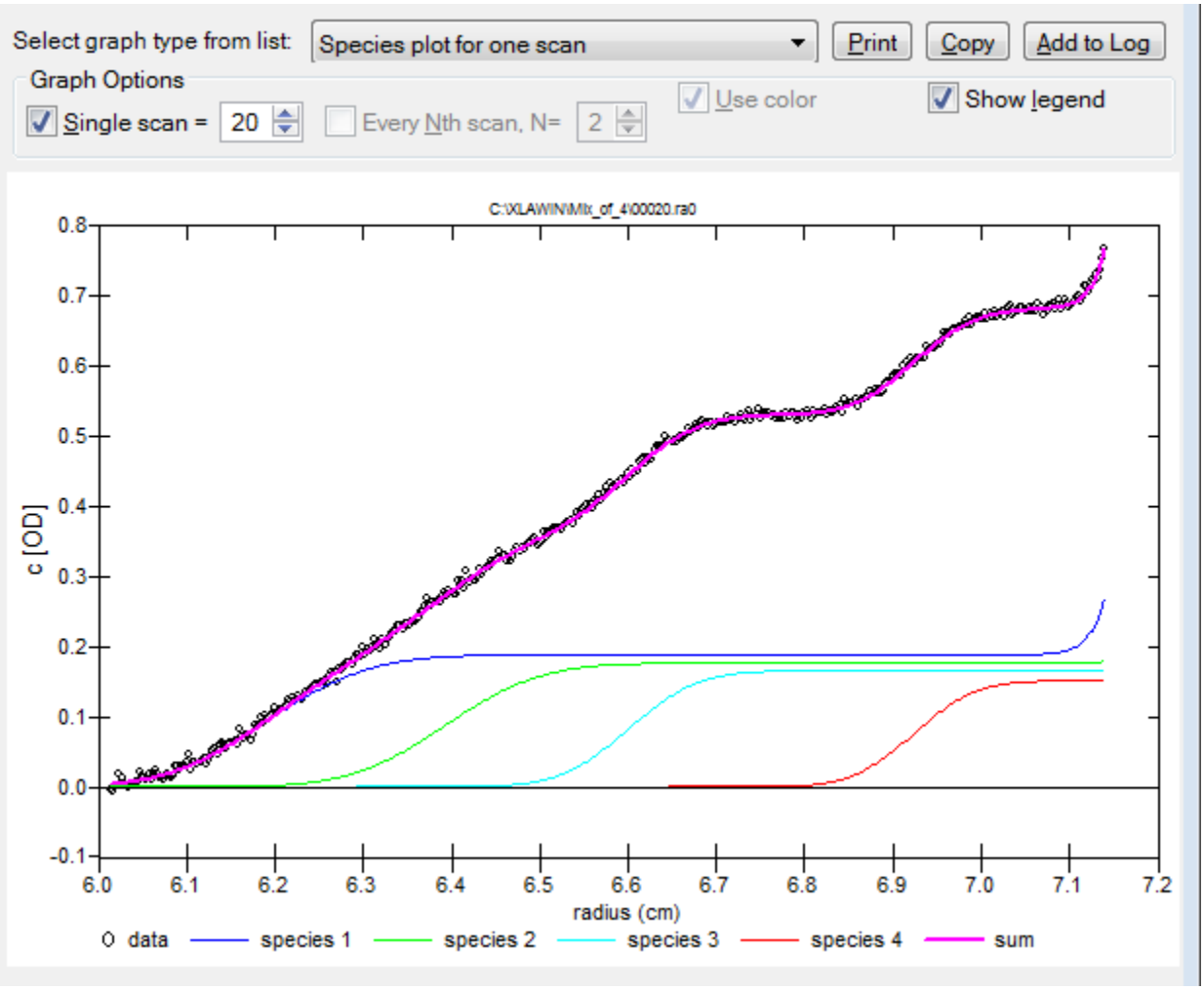
This plot shows only the actual experimental data that were fitted. When the TIN has been fitted, as in this example, it is also displayed as the black trace, and it (and the RIN if used) can optionally be removed from the raw data to get cleaner signals by checking the **Remove TIN/RIN** option.

Options are available for showing only a single one of the fitted scans, or for showing only every Nth scan.

Species plot for one scan

This plot shows the theoretical boundaries for each individual species in a multi-species fit, calculated at the time of one particular fitted scan, along with the corresponding raw data and the sum of all the species (the fitting function).

The example below comes from a simulation of a mixture of 4 fairly widely-separately species. This type of plot is useful for showing which species are still present at certain times in the run and also for judging the quality of the separation.



By default a legend showing the color key for the different traces is displayed, but if desired it can be removed by un-checking the **Show Legend** option.

Because color is normally used to distinguish the different species the **Use color** option is disabled. If you wish to print or copy/paste this graph in black and white, first right-click the graph to bring up the context menu, and then select the 'Monochrome + Symbol' option under the 'Viewing Style' sub-menu.

For this graph if the TIN has been fitted it is not explicitly displayed, but it can be removed from the raw data and sum graphs using the **Remove TIN/RIN** option.

4 How to...

4.1 How to avoid common mistakes

Problems with strange results or bad fits usually arise either from a violation of the assumptions that are inherent in this type of analysis, or instrumental problems. Below is a list of mistakes and pitfalls to avoid.

1. Wrong assumptions

It is essential to keep in mind that fitting to a sum of species requires certain assumptions about the nature of the sample and the way in which the experiment is conducted:

1. the sample has no significant concentration dependence of *s* or *D* values
2. if there are multiple species, they do not interact or inter-convert (or if they do inter-convert that process occurs slowly compared to the time of the velocity run)
3. solution non-ideality is minimal
4. the rotor is directly and rapidly accelerated to run speed
    - With regard to the last assumption regarding rotor acceleration, it is not possible to obtain correct diffusion coefficients or molecular masses for experiments where the rotor speed is altered after significant sedimentation has occurred (although the sedimentation coefficients would still be correct). Thus it is not a good idea to start the rotor spinning and then wait for the temperature to equilibrate. See **How (and why) to equilibrate the temperature (Section 4.2.1)**.
    - However, in most cases it is quite acceptable to pause at low speed (~3K) to take a quick scan to check that the samples are loaded properly. The exception to this would be for samples with very high *s* values that will sediment significantly even at low speed.

2. Insufficient sample volume

A common cause of poor resolution of multiple species in an analysis is loading too little sample into the cell (too little column height). A longer sedimentation distance improves the physical separation between species having different sedimentation coefficients, and also reduces the relative influence of diffusion on each species. It is recommended that you load enough sample to give a meniscus position <= 6.0 cm (about 430 microliters of sample in a Beckman 12 mm charcoal-epon cell).

3. Cell leakage

Another common cause of bad fits and strange results is slow sample leakage, which is easily detected by shifts of the meniscus position to higher radii over the course of the run. Slow sample leakage usually results from evaporation via a leak at the red cell plug gasket rather than a liquid leak through the window-centerpiece interface.

4. Convection caused by poor cell alignment or quality

Poor alignment of the cell housing in the rotor, a damaged key slot in the centerpiece, or a warped center rib section of the centerpiece can cause the sedimenting macromolecules to collide with the centerpiece walls. This in turn causes the concentration to build up at the wall, which increases the local solution density, which can trigger convection within the cell. Scratches or rough spots on the centerpieces walls can also trigger convection. All methods of SV data analysis assume all transport arises from sedimentation, not from convection! Convection is a disaster, no matter how you choose to analyze the data.

It is critical to use high-quality centerpieces and to align them carefully so the centerpiece walls are exactly parallel to the centrifugal force, especially if you want to measure low levels of minor components.

Convection usually does not produce any clear symptoms in the raw data. Severe convection may cause the apparent position of the meniscus to move far from its apparent optical position if the meniscus position is treated as a fitting parameter, but this is not always a clear diagnostic of convection.

5. Wavelength shifts

With absorbance data you really cannot take data at different wavelengths during the same run.

- Because the monochromator never returns to exactly the same wavelength, if you use different wavelengths then the apparent concentration will vary from scan to scan (because the extinction coefficient will be slightly different as the wavelength varies +/- 1-2 nm from scan to scan). This effect can most easily be seen as a jumping up and down of the plateau region from scan to scan. If you need data at different wavelengths, it is highly recommended that you do each wavelength as a separate run.

6. Drifting sample temperature

Another instrumental problem that can cause bad fits and poor or inconsistent results is changes in sample temperature during the run.

- It is important to thoroughly equilibrate the rotor temperature before starting the run (and remember that you are not really reading the rotor temperature until the vacuum drops below 100 microns!). See **How (and why) to equilibrate the temperature (Section 4.2.1)**.
- A plot of the temperatures recorded during the run can be obtained from the **Graph display options tab (Section 3.2.2.1)** of the **Set meniscus & data region page (Section 3.2.2)**.

7. Interference from buffer components

Especially for interference scans, you may see systematic deviations between the data and the fit in the region of the cell near the meniscus resulting from redistribution of salts or other low molecular weight components in the cell region near the meniscus.

- To minimize these problems, the sample and reference column heights can be matched exactly if you use a synthetic-boundary centerpiece (or better still the Walter Stafford meniscus-matching centerpiece design available through **Spin Analytical (http://spinanalytical.com/)**) and follow this procedure:
  1. Fill the cell as you would a conventional cell.
  2. Bring the rotor up to about 10K rpm to force sufficient solvent to cross through the capillary to equalize the volumes in both channels (you can see this happen on the Raptor fringe display).
  3. Stop the run, remove the rotor, and turn it over several times to mix the sample channel. This procedure gives a slight dilution of the sample.
  4. Reload the rotor, and re-equilibrate the temperature. When you come up to running speed the reference and sample meniscus will be exactly matched.

4.2 How to optimize your experiments

4.2.1 How (and why) to equilibrate the rotor temperature

Temperature is the key to obtaining reproducible and accurate sedimentation coefficients!

The "raw" sedimentation coefficient will typically increase by 2.4% per degree (due primarily to the large temperature dependence of the viscosity of water). Since sedimentation coefficients can be measured with a precision of ±0.2% or better, that means temperature control to ±0.1 degree is necessary to avoid compromising the accuracy of the *s(20,w)* values.

What is required for good temperature control?

1. Sufficient time to achieve temperature equilibration
2. Vacuum below 5 microns to avoid frictional heating of the rotor

Reduce the temperature equilibration time by putting the rotor and monochromator into the chamber to equilibrate, under vacuum, while you are cleaning, assembling and filling cells.

How can I tell if the temperature was stable during my run?

A plot of the temperatures recorded during the run can be obtained from the **Graph display options tab (Section 3.2.2.1)** of the **Set meniscus & data region page (Section 3.2.2)**.

However, the wrong method of equilibrating temperature will also cause problems!

Many people have been taught to bring the rotor to 3000 rpm and wait for temperature equilibrium before going to the final rotor speed. This approach will cause problems for nearly all analysis methods that attempt to derive diffusion coefficients or masses from velocity data (including this one)!

- To handle diffusion these methods must assume instantaneous acceleration to the final rotor speed. If any significant amount of sedimentation occurs at low speed, the derived diffusion coefficients and masses will be compromised (but a quick scan at 3000 rpm to check for proper cell loading should not be a problem for species <~ 10 MDa)

The trick to proper temperature equilibration: "start" a run at zero rotor speed

This will turn on the diffusion pump without spinning the rotor.

- Remember, the instrument only turns on the infrared thermometer to read the actual rotor temperature when the vacuum drops below 100 microns, so it is essential to get the diffusion pump working to get proper temperature equilibration.
- Setting the speed to zero can only be done from the instrument control panel, not from the computer through the control software

  1. first push the 'Speed' button then push '0' and 'Enter'
  2. then hit 'Enter', then 'Start'

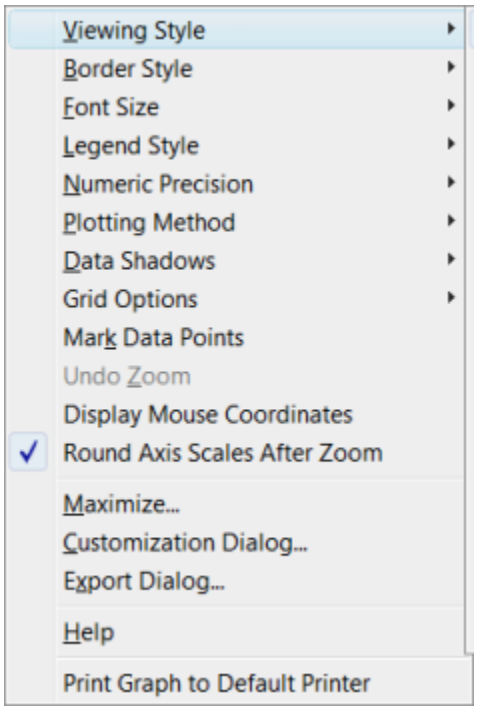
Caution! This trick creates a small problem:

If you need to release the vacuum when "running" at a speed of zero the instrument fails to force you to wait for the diffusion pump to cool (a bug in the instrument firmware). If you need to release the vacuum, hit 'Stop' and then wait ~5 minutes for the diffusion pump to cool before you push the 'Vacuum' button. Failure to cool the diffusion pump may get pump oil on your optics, including the diffraction grating in the monochromator (~$20,000 to replace).

4.3 How to work with graphs

4.3.1 How to print graphs

All graphs can be quickly printed by right-clicking on the graph to bring up this context menu and selecting the bottom 'Print Graph to Default Printer' entry.



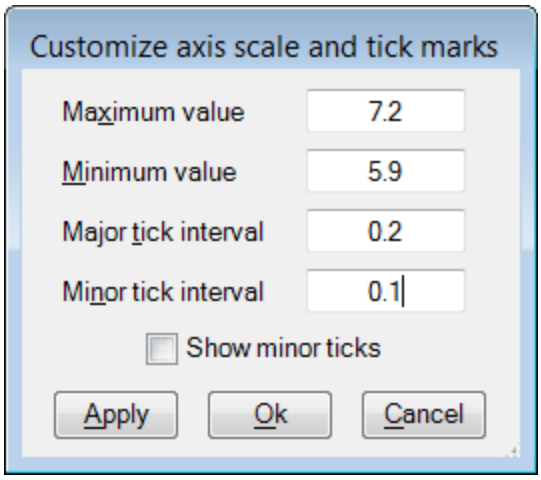
That choice will print a full-page graph in landscape orientation to the current default printer (the Windows default printer, not the printer specified for analysis reports from the **Options dialog (Section 3.1.1.4.3)**).

If you want to print to a specific printer, print a graph of a specific size, or control whether the graph is printed with color or in monochrome, you can do this by selecting the Export Dialog... menu entry.

4.3.2 How to customize graphs

'Click and edit' customizations

The axes, axis titles, and graph titles are hotspots. When passing over these hot spots the cursor changes into a hand with pointing finger; clicking will bring up a small dialog box for editing and customizing those features.

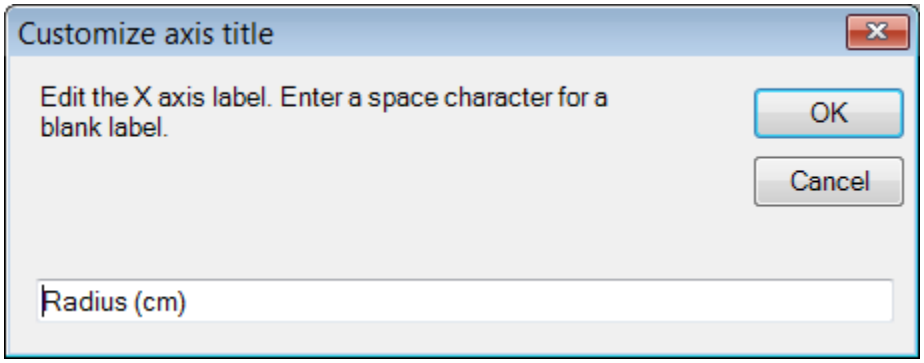


Clicking on an axis brings up the dialog shown above, which allows alterations of the scale range (minimum and maximum), the intervals between major and minor tick marks, and whether minor ticks are displayed (note that the graphs generated by DCDT+ normally do not show minor tick marks). If the graph is currently zoomed then this dialog will change the axis of the zoomed graph (and won't alter what is displayed when the graph is un-zoomed with the 'z' key).

The major tick marks (the ones which can have their values labeled) are always positioned starting from zero, not necessarily from the start of the axis. Thus in the example above, even though the X axis of this raw data plot starts at 5.9 cm, because that value is not divisible by 0.2 the labeled major tick marks will be at 6.0, 6.2, 6.4... rather than at 5.9, 6.1, 6.3...

TIP: Sometimes a tick label which should be present at the beginning or end of an axis does not display due to round-off error. When this happens you can get the label to display by increasing the axis maximum (or decreasing the minimum) a tiny amount (perhaps 1 part in 100,000), which makes that tick mark clearly within the bounds of the graph. Doing this has no visible effect on the graph scaling but solves the tick label display issue.

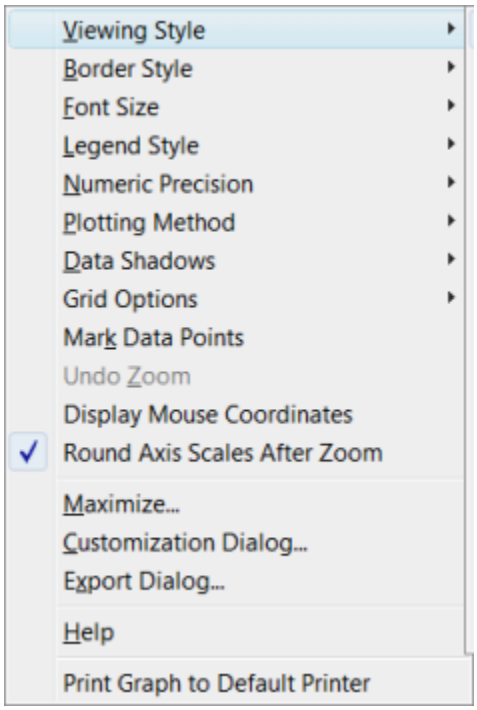
Clicking on an axis title or the graph title or subtitle (both located above the graph) will bring up small text entry dialog box that allows editing of the corresponding axis label, similar to that shown below.



Note that a completely empty entry is interpreted as the 'Cancel' command, so if you want a blank label simply enter a space character. These graph title and sub-title can also be changed through the Customization Dialog discussed below.

Right-click context menu customizations

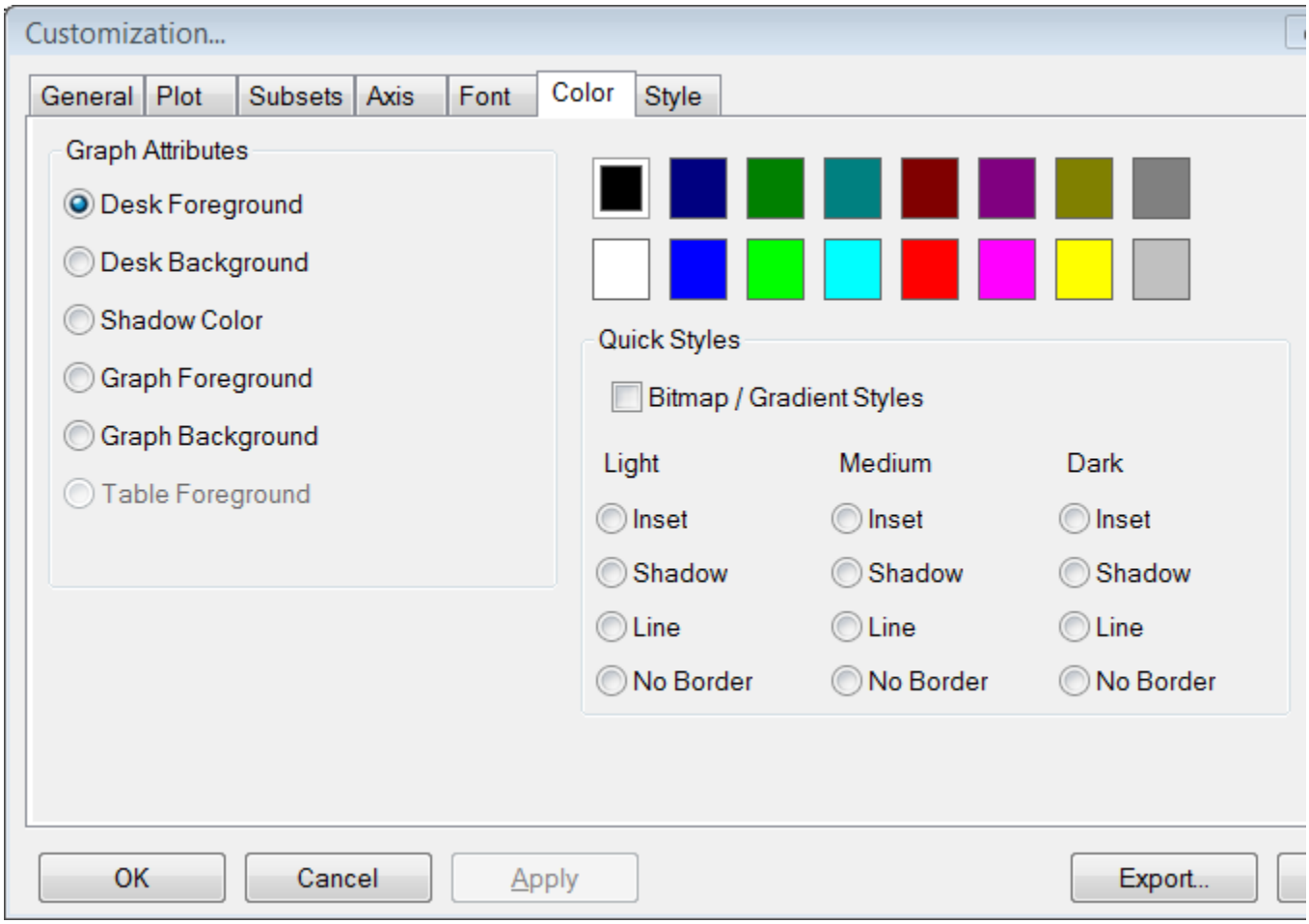
Right clicking with the mouse on any graph will bring up a context menu that provides access to numerous graph options and customization features.



The **Display Mouse Coordinates** item toggles on and off a display of the graph coordinates (X, Y values) that correspond to the current mouse position (when the mouse is positioned over the graph).

The **Round Axis Scales After Zoom** item toggles on and off whether after zooming in on a region of the graph (see **How to zoom with the mouse (Section 4.3.4)**) the graph scales are first rounded (rounding on) or instead the graph displays the exact region that was zoomed (rounding off). Rounding avoids very odd values for the scale minimum and maximum, but after zooming into a very small region it sometimes leads to a range that is much larger than you intended to display.

The **Customization Dialog** item brings up a tabbed dialog (shown below) that allows you to alter features such as line/point styles and colors, the plotting method for individual data sets (points, lines, *etc.*), legend display, legend position, graph borders, and graph background colors. This customization dialog also provides a second means of customizing axis scaling (but not tick marks, as can be done by clicking directly on the axis) and graph titles (but not axis titles).



The **Export Dialog** context menu item brings up a dialog that allows printing the graph, exporting the image to a file or the Clipboard, and even exporting the data that creates the graph. Note however that if you want to export data, this is usually done more easily using the **Edit menu (Section 3.1.1.2)**.

> For graphs that contain markers such as the vertical lines to indicate the meniscus position or regions of interest, if you don't want those markers to appear in the exported graph first uncheck the **Show Annotations** menu item before you export the graph.

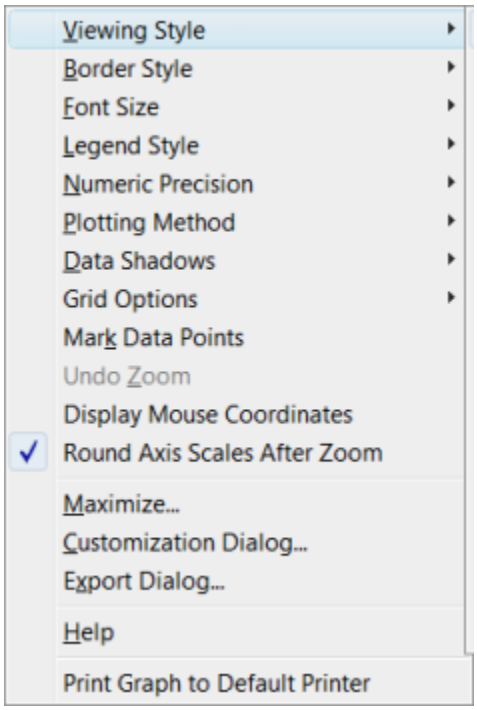
The **Help** menu item brings up a small graphing-specific Help file that discusses some of these options in more detail.

The **Print Graph to Default Printer** item sends a full-page graph in landscape orientation to the current default printer (the Windows default printer, not the printer specified for analysis reports from the **Options dialog (Section 3.1.1.4.3)**). If you want to print to a specific printer, print a graph of a specific size, or control whether the graph is printed with color or in monochrome, you can do this by selecting the **Export Dialog...** menu entry.

Most of these context-menu customization features are implemented by the **Gigasoft ProEssentials (http://gigasoft.com/)** graphing library used by this program.

4.3.3 How to export graph images or data to other programs

Right click with the mouse on any graph to bring up a context menu that provides access to various graph options and customization features, including the **Export dialog**.



The **Export dialog...** menu item brings up a dialog box that allows

1. exporting the image to a file or the Clipboard
2. exporting the data that creates the graph
3. printing the graph to any available printer

With regard to exporting data values, note that the data contained in many of the graphs can be more easily copied to the Clipboard using the **Edit menu (Section 3.1.1.2)**.

Exported image types supported include Windows metafile, Windows bitmap (BMP), JPEG, and portable network graphics (PNG).

4.3.4 How to zoom with the mouse

To "zoom in" (enlarge) a section of the graph you can click the left mouse button at the corner of a region within the axes of an existing graph and then drag (move while holding the left mouse button) until you reach the opposite corner of the region you wish to be expanded. As you drag, an inverted area will indicate the region being marked. When you stop dragging (by releasing the mouse button) the axes will be rescaled to show you this region.

If the region you have defined is very small, the zoom is not enabled, as indicated by the cursor shape.

You can undo this zoom and return to the previous scales for the axes by entering a **Z** or **z** from the keyboard, or right-clicking on the graph and using the 'Undo Zoom' entry on the pop-up menu.

Note that by default the axis scales will be rounded to even values after you zoom. Sometimes this behavior is undesirable, and it can also prevent you from zooming to very high magnification. This scale rounding behavior can be turned off by right-clicking on the graph and selecting the 'Round Axis Scales After Zoom' entry to toggle it off.

4.4 How to remove bad data points from scan files

Generally it is not necessary to manually remove individual "bad" data points from a scan file, because those can be temporarily removed by the despiking routine (see **Load raw scans - Data loading options (Section 3.2.1.7)**). If you do want to edit a scan file to remove bad data (perhaps for use by other software), methods for doing so are discussed below.

If you do need to manually remove bad data point(s), this is very easy to do:

1. Go to the **Load raw scans page (Section 3.2.1)** and be sure the **Despiking option (Section 3.2.1.7)** is turned off so that all data points are visible.
2. Advance to the **Set meniscus & data region page (Section 3.2.2)** to graph the raw data, and select the **Graph display options tab (Section 3.2.2.1)**.
3. Push the **Identify scan and data point button** to display every data point as a clickable object. (Generating this display can take a fair amount of time, so be patient.)
4. Click on the bad data point and its *x*,*y* coordinates and scan file name will be displayed. (After you click a point it can take a significant time for the third-party graphing package to identify that point, so again be patient.)
5. Click on this same point a second time and you will be asked to confirm that you want to permanently remove that point, and then asked whether you want to remove it in an irreversible or reversible manner (explained below). If you choose to make an irreversible deletion, a back-up copy of the un-altered scan file will be created (by adding an extra .bak extension on the scan file name).

Although the easiest way to remove bad points is the one described above, from the **Load raw scans page (Section 3.2.1)** you can also right-click on a scan file name and then automatically bring that scan into Notepad for editing. If you do edit scan files using the right-click pop-up menu from the **Load raw scans page (Section 3.2.1)**, the program will automatically reload that edited scan file.

Alternatives for removing bad points from scan files:

Although it possible to simply delete the data row corresponding to a bad data point from scan files, that procedure leaves no record that a bad data point was removed, and does not permit easy restoration of the original data. To address these concerns, this program implements the "University of Connecticut convention" for marking bad data points in a manner where the original information is retained so it can later be restored if needed.

- Note however that not all AUC analysis software follows this convention, and in particular Peter Schuck's SEDFIT and SEDPHAT do not, so if you want to use SVEDBERG to edit scans for those programs you must permanently delete the bad data points.

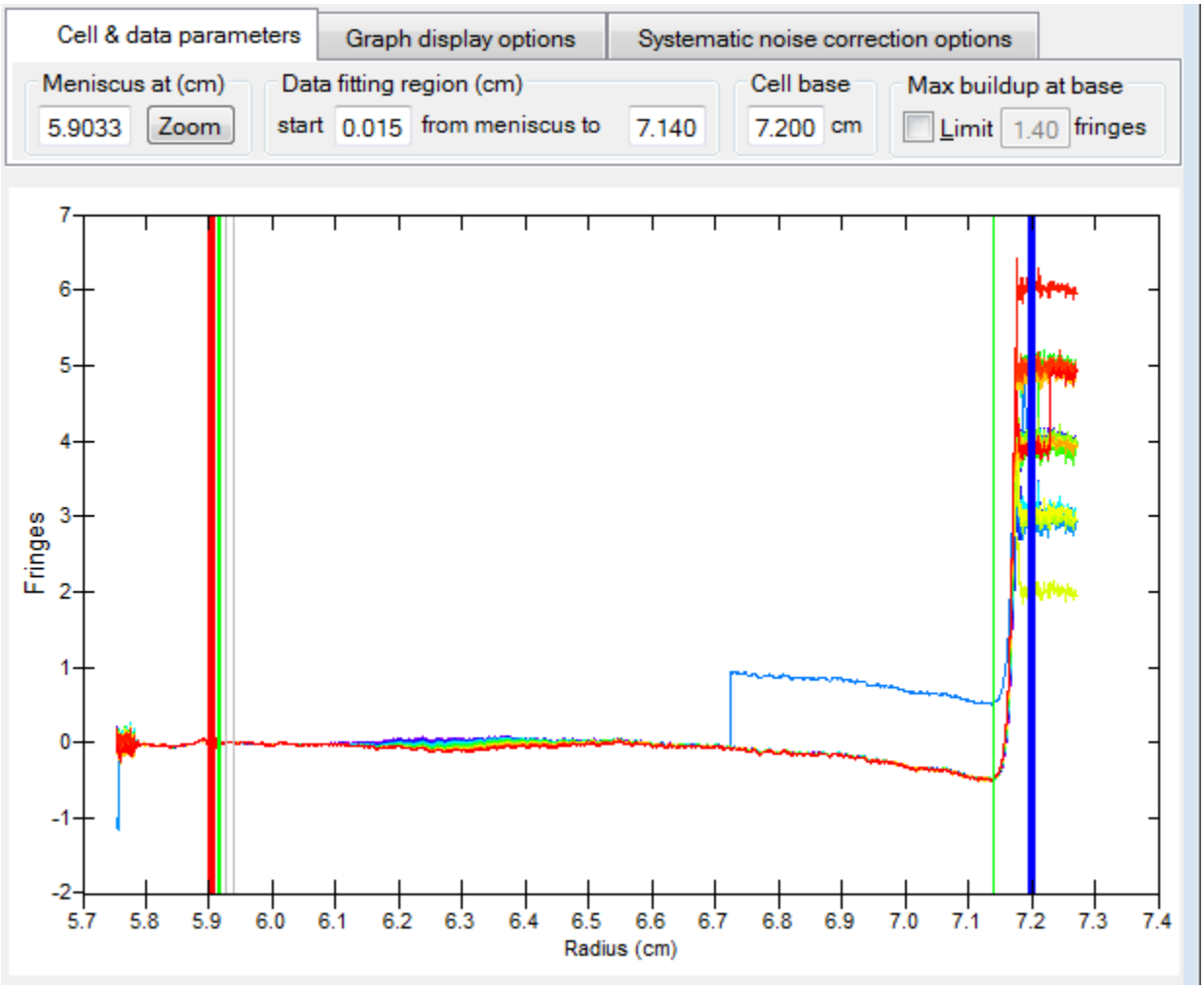
By this "UConn convention", any data point for which the value of the third column in the data is negative will be ignored when the file is read. For absorbance data, this third column normally contains either zero (when no replicates are used) or the standard deviation of the absorbance value (when multiple flashes are being averaged). For interference data, there normally is no third column entry. However, this software checks each line of data for a third column, and will ignore any data point where a third column exists and is negative.

Thus to mark a data point as "bad" the entry in the third column is set to a negative value (by adding a third column for that line with -1 as the value in the case of interference data). For absorbance data, simply changing the sign of the standard deviation allows you to later restore the original raw data if needed, without loss of information, by changing the sign back to positive again.

4.5 How to identify and remove rogue scans

The interference optical system may sometimes produce "*rogue scans*". Needless to say, you don't want to use these aberrant scans in your analysis!

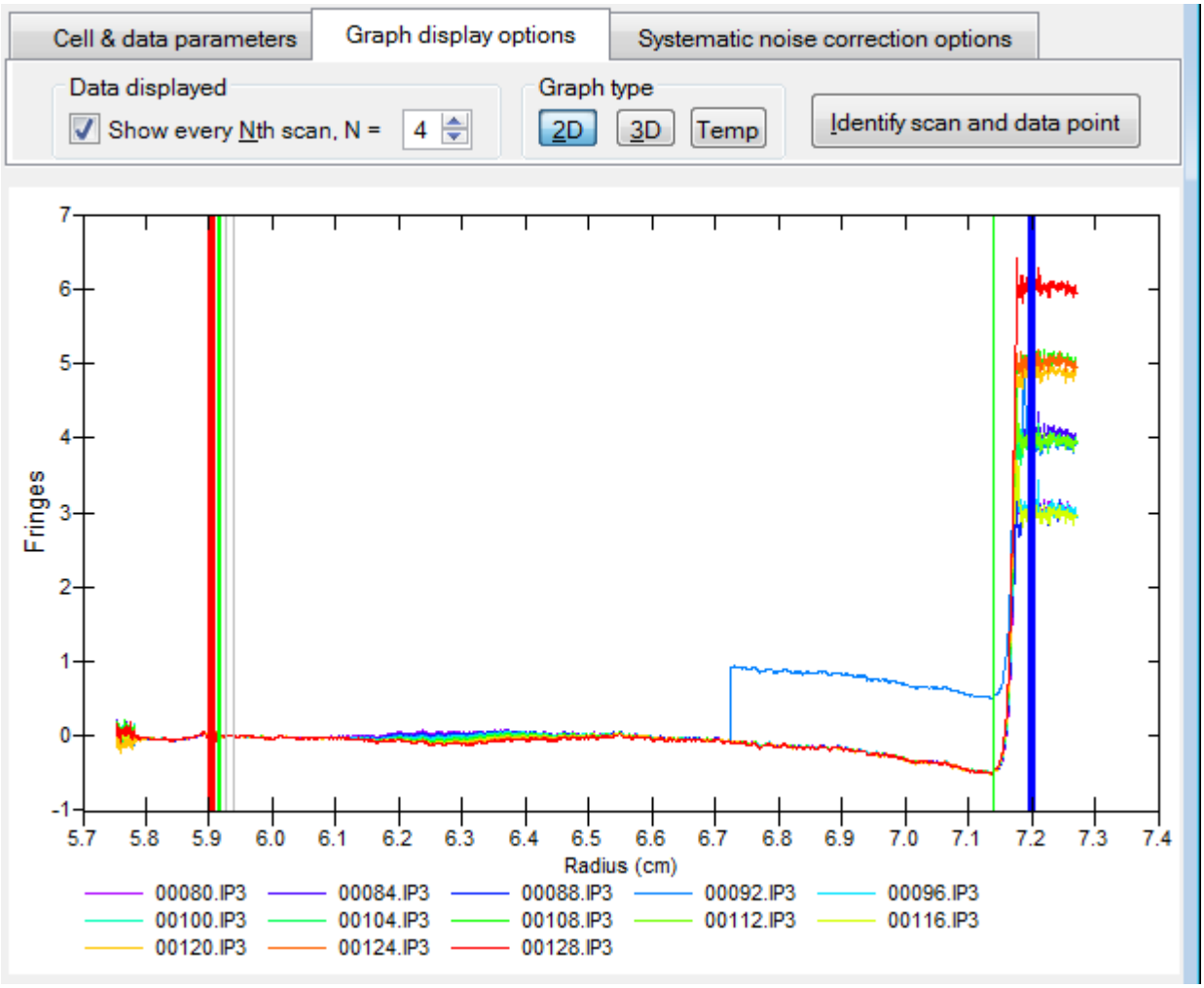
You may see such a scan when you view the scans on the **Set meniscus & data region page (Section 3.2.2)**. The image below shows an example where there is an integer fringe jump within the sample region.



Because there may be dozens or even hundreds of scans in this display, it can be difficult and frustrating to try to identify which is the bad one.

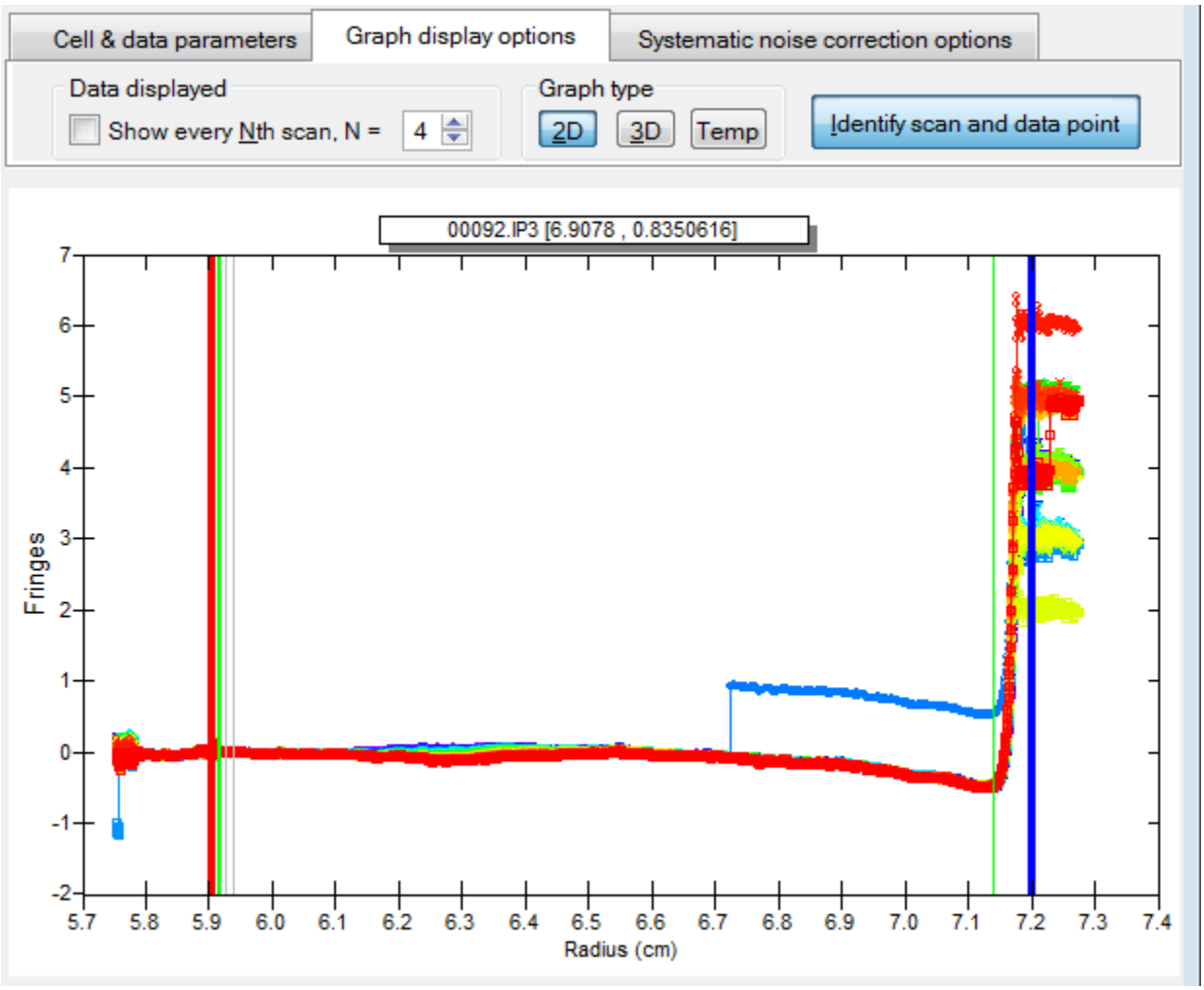
Sometimes you may be able to identify the rogue scan by reducing the number of scans displayed so the line colors become more distinctive and to make the number of scans low enough that a legend can be displayed. (By default this graph does not display a legend when there are more than 40 scans shown, as is true above, because the legend occupies too much space.) To reduce the number of scans displayed, select the Graph display options tab, click the **Show every Nth scan** option, and increase **N** as needed until a legend is displayed and the rogue scan is visible. Note that the weakness of this approach is that not all scans are shown and thus the bad one may or may not show as you change the value of **N** (and in some cases it will not be visible for any value of **N**).

In the example below it is now possible to distinguish from its color that it is scan 92 (file 00092.ip3) that is

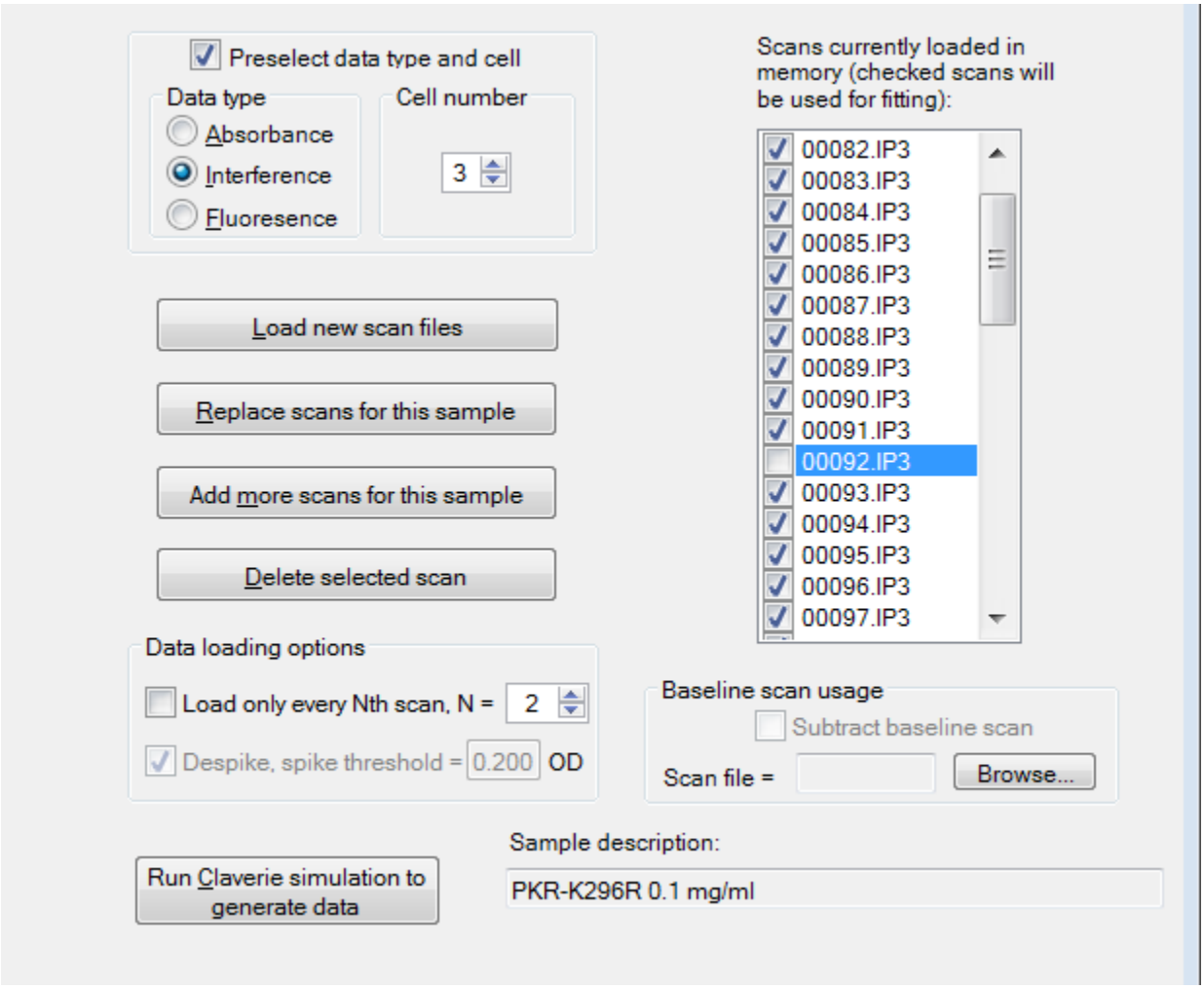


bad.

However an alternate procedure that always works is to turn off **Show every Nth scan** and instead push the **Identify scan and data point button**. This turns every point in every scan into a clickable object, and places a window above the graph that will display the scan file name and X,Y data values for any point that is clicked. (The drawback of this approach is that generating this display can take a fair amount of time, so be patient.) Now if you click anywhere along the right portion of the rogue scan (the easily-identified region to the right of the fringe jump), then as shown below the window above the graph identifies that the rogue scan is number 92. (After you click a point it can take a significant time for the third-party graphing package to identify that point, so again be patient.)



Once the rogue scan is identified, use the **Back button (Section 3.1.3.1)** to return to the **Load raw scans page (Section 3.2.1)**. Click on the file name of the rogue scan within the **Scans currently loaded in memory (Section 3.2.1.6)** list box, and then click again to un-check that scan so it will be ignored, as shown below, and then continue with your analysis.



Alternatively you can first select the bad scan in the **Scans currently loaded in memory (Section 3.2.1.6)** list box and then push the **Delete selected scan (Section 3.2.1.5)** button to completely delete the rogue scan from memory, and then continue with your analysis.

4.6 How to minimize the effects of a sloping plateau

Why is a sloping plateau (Section 5.1.5) a concern?

Since theoretically the plateau should be flat, it is impossible for the fitting functions to accurately represent data with a sloping plateau. This problem will have little impact on the returned sedimentation coefficients, but can result in significant errors in M or diffusion coefficients, and in estimating the concentrations of minor species.

How can these effects be minimized?

1. Try to reduce this optical artifact by increasing the lamp intensity
   - Have your lamp cleaned, pick a wavelength where there is more light, or reduce the sample concentration
2. Do one of the following:

1. Use the '**slope correction' fitting parameter (Section 3.2.4.6)**
2. Set the outer end of the radial range for each scan to the leading edge of the boundary, eliminating the plateau regions, by clicking the **Limit S(max) checkbox** in **Select fitting model - Data options (Section 3.2.3.4)** (by using partial boundary modeling) and setting an appropriate limit value.
3. Use the **delta-c mode (Section 3.2.3.4)**. The subtraction used in this mode tends to eliminate the slope of the plateau. This works better when the scans are closer together in time.

5 Frequently-asked questions

5.1 About fitting

5.1.1 Should I fit the meniscus position?

Definitely fitting the mensicus position will be necessary if that position is not well defined in the raw data. For fluorescence scans this is often true, and it may also be true for interference scans. This can easily be true for all optical systems when you use meniscus-matching centerpieces.

With absorbance data the sample meniscus usually produces a large positive spike. For ~60 years it has been traditional to assume that the top of this spike represents the "true" meniscus position, and indeed that is the convention followed by the **meniscus wizard (Section 6.4)**. On the other hand, this spike does represent an optical artifact (the absorbance optics are not explicitly designed to measure the meniscus position), and the shape or position of the spike could be affected by the refractive index of the solvent or the presence of surfactants at the air-liquid interface. Thus it is not clear that the top of the spike is really the correct position.

Then why not just always fit the meniscus position?

One problem with fitting the meniscus position is that if the model being fitted is wrong (it does not completely and accurately account for the behavior of the sample and all sources of systematic error) then often the fitter will compensate for the inadequacy of the model by moving the meniscus away from the "true" position. If for example your sample actually contains more species than you have assumed are present, or if those species are associating and dissociating during the experiment (the sample does not behave as a simple mixture), then in extreme cases the best-fit meniscus position may occur at a location that is physically unrealistic (clearly away from the optical distortions which mark the meniscus).

Other model deficiencies which may affect the fitted meniscus position

These effects are not accounted for by the fitting functions in SVEDBERG and thus may cause the fitted meniscus position to move away from the "true" position:

1. Convection in the cell caused by temperature gradients, scratches on channel walls, bowing of the rib between the two channels, or poor rotational alignment of the cell.
2. The presence of gradient-forming components in the buffer such as sucrose or glycerol.
3. Solution non-ideality effects (which may occur even at low concentrations for highly charged solutes at low ionic strength).
4. Using solvents that are much more compressible than water.
5. The tendency of some solutes to be attracted to the air-liquid interface and accumulate there (that is, the meniscus itself is a compartment that differs from the bulk solution).
6. Diffusion that occurs during the acceleration phase of the rotor.
7. The slow scan speed of the absorbance optics (boundary motion that occurs during an individual scan).

In other words, when the fitting model is not completely accurate, the "best fit" in the least-squares sense will not necessarily give the most accurate values for the fitting parameters, and could even give results which are physically unrealistic. While it is true that fitting the meniscus position rather than holding it fixed at a value selected manually or by the **meniscus wizard (Section 6.4)** will almost always produce a "better" fit (lower *rms residual*), note that it has never been demonstrated that for real experiments fitting the meniscus position gives results that are either more accurate or more precise.

A second important reason why fitting the meniscus position may be undesirable is that adding another fitting parameter usually reduces the precision of all the other fitting parameters (it increases the size of the confidence intervals). In some circumstances this additional fitting parameter may lead to a failure of the fit to converge.

Overall, whether it is better to fit the meniscus position or not may depend in part on which information is most important to you. The values for sedimentation coefficients are most affected by the meniscus position, and if the fitted position is clearly moving away from where the raw data suggest the meniscus is located then the accuracy of the returned sedimentation coefficients seems questionable.

5.1.2 Why should I limit the buildup at the cell base?

Although the **Behlke function with base (Section 3.2.3.3)** fitting function provides an accurate theoretical model for the buildup of solutes at the cell base, it is nonetheless often a good idea to at least exclude the portion of the raw data where the buildup reaches high concentrations by using the **Limit maximum buildup at cell base (Section 3.2.2.8)** option, and it may be a good idea to exclude even more of the data near the cell base (by changing the **Good data end at radius = (Section 3.2.2.5)** value).

The reasons why the theoretical model will often fail to provide a good fit for the high concentration regions near the cell base include:

1. The OD exceeds the linearity of the absorbance optics.
2. Some of the species present in the sample may irreversibly aggregate or precipitate as their concentration builds up to high levels at the base. This means that the theoretical fitting function cannot possibly correctly account for the concentration distribution of all species near the cell base, and to compensate for that the effective position of the cell base may need to move away from the "true" position.
3. The concentration at the cell base often gets high enough that solution non-ideality effects are quite significant. Because those effects are not included in the fitting models, the cell base position which will provide the best fit may not be the "true" position.
4. The true base position is actually not well defined for most velocity cells because the bottom of the cell is flat rather than being curved along a radius.

Due to points 1-4 above it is best to avoid having your fit be strongly dependent on the details of what is happening near the base of the cell.

5.1.3 Should I fit the cell base position?

If the region of the cell that is being fitted is being significantly influenced by the accumulation of solutes then the exact location of the base of the cell may be important, and in general when that is true it is necessary to treat this position as a fitting parameter.

Although in principle it should be possible to measure the "true" position of the cell base either by examining the raw interference fringes (for an XL-I) and/or by requesting an absorbance intensity scan, that is not necessarily the best choice for the base fitting parameter, because:

1. The true base position is actually not well defined for most velocity cells because the bottom of the cell is flat rather than being curved along a radius.
2. Some of the species present in the sample may irreversibly aggregate or precipitate as their concentration builds up to high levels at the base. This means that the theoretical fitting function cannot possibly correctly account for the concentration distribution of all species near the cell base, and to compensate for that the effective position of the cell base may need to move away from the "true" position.

3. The concentration at the cell base often gets high enough that solution non-ideality effects are quite significant. Because those effects are not included in the fitting models, the cell base position which will provide the best fit may not be the "true" position.

Due to points 1-3 above it is best to avoid having your fit be strongly dependent on the details of what is happening near the base of the cell. Thus you may want to exclude a large part of that region from the fitting, or at least to use the **Limit maximum buildup at cell base (Section 3.2.2.8)** option.

5.1.4 Which method of systematic noise removal should I use?

5.1.4.1 Introduction

Velocity scans usually contain systematic baseline deviations (zero concentration does not produce zero signal), and also may include vertical displacements from one scan to the next ('jitter' or radially-invariant noise). Therefore it is important to either remove this systematic noise before the data is fitted, or to explicitly include the systematic noise as part of the model to be fitted.

There are several alternate approaches to handling this systematic noise, which each differ mathematically, statistically, and even philosophically. The good news is that SVEDBERG implements all of the current approaches, and this makes it easy to try all of them and see how the results may differ. The bad news is that this means that ultimately you the analyst must decide which one is 'best' and which results to report. Therefore the purpose of this topic is to discuss the advantages and drawbacks of each approach to help guide that decision.

Note that this topic remains fairly controversial in the field (one reason most software packages only offer a subset of the options).

5.1.4.2 Removal or fitting of baseline offsets or profiles

The three basic options for handling baseline offsets (a constant value at all radii) or baseline profiles (time-invariant noise or TIN) are:

- record an experimental baseline scan, and **subtract that from each scan (Section 3.2.1.8)** to remove this systematic noise prior to fitting
- subtract the scans in pairs ('scan differencing' or '**delta-c mode (Section 3.2.3.4)**'), which removes the time-invariant noise by simple arithmetic
- explicitly include in the fitting model either a **simple zero offset or an unknown baseline profile (Section 3.2.3.5)** (a set of fitting parameters giving the zero concentration reading at each radius value) and determine these parameters as part of the least-squares fitting procedure

The following four expandable sections describe the advantages and drawbacks of each of these approaches.

Subtract an experimental baseline

In principle it is always best to use an experimental baseline. (The issue of how to record one is discussed **here (Section 3.2.1.8)**.) Doing so avoids the decrease in signal/noise ratio that is inherent with scan differencing, and avoids introducing extra unknown fitting parameters (a baseline offset or TIN). However there are two potential drawbacks of using an experimental baseline:

1. photometric noise in the baseline scan will increase the total noise in the data being fitted;
2. if the baseline scan is not accurate then it could introduce systematic errors.

The types of things which could make the baseline scan inaccurate depend on how it was recorded. If the sample was rinsed out, this could potentially rinse away sample molecules that were loosely stuck to the windows, and the cell position in the rotor may not match what is was during the main run. Leaving the cell in and trying to fully pellet the sample could introduce errors if the sample or reference buffer accidentally contain some small molecules which absorb at the measuring wavelength and the concentrations and meniscus

positions are not perfectly matched.

- **Scan differencing (delta-*c* mode)**

The major advantages of scan differencing over fitting the TIN are:

1. Scan differencing requires no additional fitting parameters, whereas fitting the TIN requires hundreds of extra parameters
2. Scan differencing is model-independent. The removal of time-invariant noise is essentially axiomatic---by definition any noise that is time-invariant will be completely and exactly removed by simple arithmetic. Further, the noise removal occurs separately from, and prior to, the choice of any model to explain the changes in signal with time. Unlike with fitting of the TIN, the net time-dependent signals do not change when the sedimentation coefficient, masses, or concentrations change.
3. Unlike fitting the TIN, scan differencing does not require that you include scans covering the full range of boundary movement (all species must move completely through and out of the regions of the cell that are included in the fit).

The major drawbacks of scan differencing relative to fitting a baseline offset or TIN are:

1. The differenced data have a lower signal/noise ratio than the original scans, and the total number of data points being fitted is also reduced by a factor of 2. Thus we expect that scan differencing will lead to some degradation of the precision of the fitted parameters, but whether this is significant depends on the signal/noise ratio of the raw data and how much precision is needed.
2. Differencing mixes together data from different times in the run, which makes interpretation of patterns in the residuals more difficult.
3. Scan differencing does not yield an actual baseline profile that could be subtracted from the raw data or compared to an experimental baseline.

- **Simple baseline offset**

The advantage of a simple offset over fitting the full TIN is that it requires only one unknown fitting parameter rather than hundreds. Its advantages over scan differencing are that it does not change the signal/noise ratio of the data being fitted, and does not mix together data from different times in the run.

The primary disadvantage of using a simple offset is that it is seldom an accurate description of the time-invariant systematic noise. An offset is at best only a moderately good first-order description of the true baseline for absorbance or fluorescence scans. A further disadvantage is that each additional fitting parameter tends to lower the precision of the other fitting parameters, but that is usually a relatively minor drawback (especially compared to fitting the entire TIN profile).

- **Fitted TIN**

Clearly fitting the TIN profile provides much more complete and accurate removal of time-invariant noise than a simple baseline offset.

The major advantages of fitting the TIN over scan differencing are:

1. The raw data are directly fitted, and from a statistical point of view the systematic noise and the sedimentation model are being treated equivalently and simultaneously (globally).
2. The determined baseline profile can be compared to an experimental estimate (if available), and also checked for unreasonable features such as sudden changes in slope in certain regions of the cell, as a "reality check" to flag whether there may be some problem with the model or the experimental data itself.
3. The fitted TIN represents an average derived from all the scans, and thus has only a small random noise component. Scan differencing actually increases the random photometric noise in the difference data by approximately the square root of 2.

The major drawbacks of fitting the TIN are:

1. The computed TIN is model dependent. If the fitting model is correct, and if you obtain a good fit (no systematic deviations) then this fitting procedure will give the correct baseline. However when the model

is wrong (wrong number of species, the sample does not behave as a simple mixture, or non-ideality effects are significant), or if the fitting functions do not perfectly model the data, then the fitted baseline tries to compensate for the model's inadequacies by distorting the baseline. This can lead to inaccurate results, and potentially could even lead to bad outcomes such as false species.

2. Because the computed TIN changes as the sedimentation coefficient, mass/diffusion, and concentration parameters move away from the best-fit values the increase in *chi squared* is reduced (the error surface is less strongly curved). This leads to lower parameter precision (larger confidence regions) and can lead to poor convergence of the fit. Particularly for fits using **partial-boundary modeling (PBM) (Section 3.2.3.4)** this loss of parameter precision can be substantial (see **ref. 7 (Section 7.7)**).
3. Because the fitted TIN always forces the average *residual* at each radius to be zero, it tends to 'hide' defects in the fitting model (it makes it more difficult to distinguish a bad fit).
4. Correctly evaluating the TIN requires that you include scans covering the full range of boundary movement (all species must move completely through and out of the regions of the cell that are included in the fit). Fitting only a portion of the run risks getting results which are incorrect or not meaningful.

5.1.4.3 Removal of radially-invariant noise (RIN, jitter)

The two basic options for removing signal offsets that vary from one scan to the next are:

1. **Forced alignment** over a designated radial region. This procedure is based on the principle that there are regions in the cell where nothing is sedimenting and therefore all scans should align vertically in that region (the average signals across that region should be the same). Therefore any difference between the first scan and subsequent scans is removed by subtracting an offset (the RIN value for that scan). One such region is the air-air space; after the meniscus is cleared the region just to the right of the meniscus may also be a suitable region. The alignment region is designated on the **Set meniscus & data region page (Section 3.2.2)** under **Jitter removal options (Section 3.2.2.2.1)**
2. **Compute the best-fit RIN for each scan**. In this approach the RIN values for the second through last fitted scans (their offset relative to the first scan) are included as fitting parameters in the least-squares fitting procedure. This option can be chosen on the **Select fitting model page (Section 3.2.3)** under **Baseline fitting options (Section 3.2.3.5)**.

The following two expandable sections discuss the advantages and drawbacks of each of these approaches.

Forced alignment

The advantages of aligning the fringes over a designated region in the cell *vs.* using a fitted RIN are:

1. This option introduces no additional fitting parameters.
2. Forced alignment is model independent. Further, the RIN removal occurs separately from, and prior to, the choice of any model to explain the changes in signal with time. Unlike with fitting of the RIN, the net time-dependent signals do not change when the sedimentation coefficient, masses, or concentrations change.

The drawbacks of forced alignment relative to fitting the RIN are:

1. The manual choice of the alignment region might unintentionally introduce bias into the analysis.
2. If there actually is some change over time in the chosen alignment region this could lead to incorrect results.
   For example, the region chosen is to the right of the meniscus, and the macromolecules have actually cleared that region, but there is some change there due to sedimentation of small molecules.
3. The alignment region usually contains only a fairly small number of data points so there is relatively little averaging of photometric noise.

Fitted RIN

The advantages of using a fitted RIN *vs.* a forced alignment are:

1. The entire fitted radial range is used in determining the RIN for each scan, so there is extensive signal averaging.

2. The RIN is determined based on the fitted range, not some position elsewhere in the cell where the noise could be slightly different (if the systematic noise is not truly radially invariant).
3. There is no potential for unintentional bias through manual selection of an alignment region.

The drawbacks of fitting the RIN relative to forced alignment are:

1. Fitting the RIN introduces *N* - 1 additional fitting parameters, where *N* is the number of scans in the fit.
2. Because the computed RIN changes as the sedimentation coefficient, mass/diffusion, and concentration parameters move away from the best-fit values the increase in *chi squared* is reduced (the error surface is less strongly curved). This leads to lower parameter precision (larger confidence regions) and can lead to poor convergence of the fit. Particularly for fits using **partial-boundary modeling (PBM) (Section 3.2.3.4)** this loss of parameter precision could be significant (see **ref. 7 (Section 7.7)**).
3. The fitted RIN is model dependent. If the fitting model is correct, and if you obtain a good fit (no systematic deviations) then this fitting procedure will give the correct RIN. However when the model is wrong (wrong number of species, the sample does not behave as a simple mixture, or non-ideality effects are significant), or if the fitting functions do not perfectly model the data, then the fitted RIN tries to compensate for the model's inadequacies by distorting the relative offsets between scans.

5.1.5 What is a 'sloping plateau'?

Many XL-A/I instruments exhibit a 'sloping plateaus' artifact in absorbance scans, at least at some times. This problem manifests itself as an apparent rise (or, rarely, a fall) in absorbance across the plateau region, even when the true concentration across that region is uniform.

The rise or fall in absorbance generally appears to be approximately linear with radius, and thus the plateau regions appear to slope uniformly upward or downward. This systematic error distorts the true shape of the boundaries, and thus has a negative impact on fitting results. The slope correction fitting parameter was introduced in an attempt to minimize these effects. Other approaches to minimizing the impact of this optical artifact are discussed here.

The exact cause of this problem is unclear, but in most cases it is probably due to variations in stray light across the cell, with more stray light usually contaminating the signals at smaller radii. The general observations consistent with this hypothesis are:

- the slope is always greater (as a fraction of the total absorbance) at higher absorbance
- this problem is usually worst when the lamp is dirty or at wavelengths where the amount of light is low
- the values of absorbance at the inner radii are usually below what is expected based on measurements with a bench-top spectrophotometer, and rise toward the correct value, consistent with excess stray light in the sample channel at smaller radii.

Another instrument problem that can cause a sloping plateau is if the wavelength varies with radius, and the current wavelength does not correspond to a maximum or minimum in the absorbance spectrum. This possibility is most likely to be seen when working at a wavelength where the absorbance rises or falls steeply with wavelength (for example, at 230 nm for proteins). The wavelength will shift with radius if the front face of the monochromator unit is not parallel with the motion of the slit assembly.

It is not always easy to distinguish between sloping plateaus and the true presence of a broad range of rapidly-sedimenting species (as might be present for a sample containing a range of aggregates or showing indefinite self-association).

Testing for sloping plateaus

A robust test for the sloping plateau problem consists of:

1. Load a 2-channel cell with sample, allow it to sit for ~1 hr to allow for possible adsorption to cell and window surfaces, and then before placing the cell into the rotor turn it over several times so the air bubble remixes it and the concentration will be uniform
2. Bring the rotor to low speed (typically 3000 rpm) and record a scan
3. Take the rotor to high speed (normal run speed for your main species) and immediately take another scan
4. Compare the plateau regions of the low and high speed scans. If the slope in the plateau is truly due to rapidly-sedimenting species, then at high speed such species will rapidly pellet and the absorbance across the plateau will be significantly lower than that in the low speed scan
5. If the plateau regions have the same slope and absorbance magnitude at low and high speeds (other than a radial shift due to expansion of the rotor), then this slope must be due to an instrument artifact rather than true sedimentation.

Note that the sloping plateau artifact can sometimes be seen even when using the Beckman-Coulter absorbance calibration solutions or a dye, either of which clearly will not sediment.

5.1.6 What does high parameter cross-correlation mean?

Parameter cross-correlation values near 1 or -1 indicate that the best-fit value for that parameter is extremely sensitive to the value for another parameter. Thus these two parameters cannot really be independently determined, and neither is truly well-defined.

Technically, the high correlation means that the increase in the sum of squared residuals that arises when a parameter is moved away from its optimum value can be almost completely compensated by simultaneously adjusting the value of the other parameter.

It is not at all uncommon for the concentration parameters in multi-species fits to be highly negatively correlated, and this is generally not a cause for concern. High cross-correlations for the other parameters is, however, usually a significant concern, and probably means that the data contain insufficient information to determine all the parameters uniquely. This condition might be relieved, for example, by reducing the number of species or by using the procedure.

The likelihood of high negative correlations between species concentrations arises because it is generally true that it will be possible to largely compensate for the effect of decreasing the concentration of one species by adding the same amount to another species.

TIP: One important way to reduce parameter cross-correlation in multi-species fits is to reduce the number of fitting parameters by constraining the ratios of molecular masses or sedimentation coefficients between species.

5.1.7 What constitutes a good fit?

Judging "good" and "bad" fits is somewhat of an art, and ultimately relies on the accumulated experience of the analyst. However, in general a good fit is one where:

- the residuals are small and randomly distributed (they are not systematically above or below zero over substantial portions of the radial range, and hence the *runs test Z* score is low)
    - the 2D grayscale/color 'bitmap' residual plot available on the **Graph fit results page (Section 3.2.7)** allows subtle non-random patterns in the residuals to be detected

- the *rms residual* is consistent with the intrinsic optical noise of the instrument
- the returned values make scientific sense based on other information you have about this sample

With regard to the last point, one important criterion is whether the apparent molecular weight implied by the *s/D* ratio is consistent with the sequence weight, values from mass spectrometry, or sedimentation equilibrium results.

For truly homogeneous, non-associating samples the apparent *M* from fitting should be within 2-5% of the "correct" value.

It is important to remember that the presence of extra species or self-association will make the boundaries broader, which makes the apparent *D* larger and therefore the apparent *M* smaller. Thus if the apparent *M* is more than ~5% below the "correct" value, it is likely that your sample is heterogeneous or self-associates. If this is due to self-association, you should see shifts in *s* values by changing sample concentration.

A second important criterion is to ask whether the results make hydrodynamic sense. Is a species sedimenting faster than is theoretically possible for its molecular mass ($f/f_0 < 1$)? Is a species sedimenting much more slowly than is reasonable for its molecular mass (f/f_0 very high)?

Causes of "bad" fits

A number of causes of bad fits are discussed under **How to avoid common mistakes (Section 4.1)**.

5.2 Does SVEDBERG utilize multi-core processors?

The SVEDBERG code has not yet been modified to enable running the computations in parallel on multiple processor 'cores', but this will hopefully be done for some future release.

In the meantime, if you have a multi-core processor you can run multiple copies of SVEDBERG (start the program more than once to create multiple 'instances'). For example it would be advantageous to have one program copy running a long computation such as calculating confidence levels using Monte Carlo simulations while you are doing new fits using another copy, letting the operating system assign different cores to the different copies.

5.3 How should I cite this program?

1. The primary citation for this program is:
    - Philo, J. S. (1997). An improved function for fitting sedimentation velocity data for low-molecular-weight solutes. *Biophys. J.* 72, 435-444.
2. The program itself should simply be cited as SVEDBERG (by convention program names are normally all capitals) by John Philo, with the version number.
3. The citation for the Behlke fitting function used by default starting with version 7.0 is:
    - Behlke, J. and Ristau, O. (2002). A new approximate whole boundary solution of the Lamm differential equation for the analysis of sedimentation velocity experiments. *Biophys. Chem.* 95, 59-68.
4. The citation for the multi-section option for scan differencing (delta-*c* mode) is:
    - Philo, J. S. (2011). Limiting the sedimentation coefficient for sedimentation velocity data analysis: Partial boundary modeling and *g(s*)* approaches revisited. *Anal. Biochem.* 412, 189-202.

6 Reference information

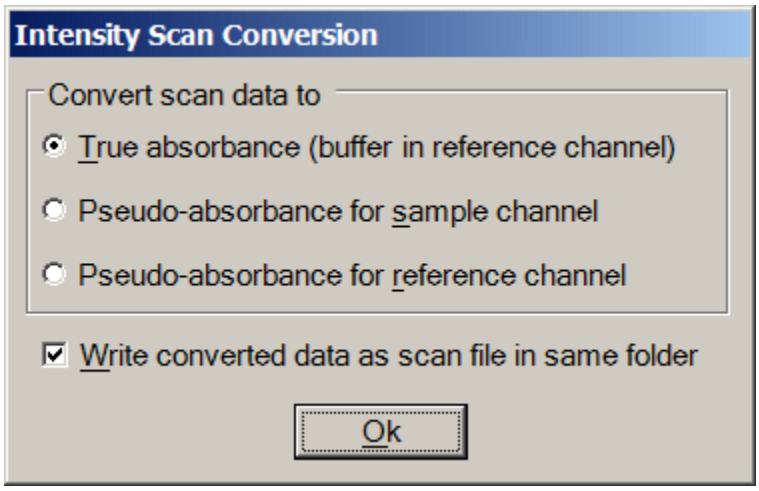
6.1 Symbols used

D	diffusion coefficient
D(20,w)	diffusion coefficient corrected to standard conditions (20 °C, water), but not extrapolated to zero concentration
M	molecular mass
R	universal gas constant
s	sedimentation coefficient
*s**	apparent sedimentation coefficient (not corrected for the effects of diffusion)
s(20,w)	sedimentation coefficient corrected to standard conditions (20 °C, water), but not extrapolated to zero concentration
ρ	solvent density
$\bar{v}$	partial specific volume
ω	angular velocity [$\omega = 2\pi \times$(RPM/60)]
$\omega^2 t$	product of angular velocity squared times elapsed time in the run (proportional to total sedimentation force)

6.2 About using radial intensity scans

In addition to normal absorbance scans, this program can also directly load radial intensity scans, including so-called **pseudo-absorbance scans** where two different samples were loaded in one cell (one in the sample channel, one in the reference channel).

When the program detects that you are trying to load intensity scans, it brings up the following dialog box for you to select how you want the data to be handled.


Intensity Scan Conversion
Convert scan data to
True absorbance (buffer in reference channel)
Pseudo-absorbance for sample channel
Pseudo-absorbance for reference channel
Write converted data as scan file in same folder
Ok

Processing and handling of pseudo-absorbance data

When you select either of the pseudo-absorbance options, the program assumes a maximum intensity value of 1000 and converts all the data points relative to that value using the formula $OD = log(1000/I)$.

When the "Write converted data as scan file in same folder" check box is checked (which is true by default), as the intensity scan files are read the converted absorbance or pseudo-absorbance values will be written out as a normal radial absorbance scan file, into the same folder from which the intensity scans are being read. For pseudo-absorbance data, the file name will be preceded by an "S" or "R" to indicate whether this data is from the sample or reference channel, respectively. That is, pseudo-absorbance scan file 00001.RI1 will generate new files S00001.RA1 or R00001.RA1.

If these converted files are later read by the program, the only way it can distinguish that these were pseudo-absorbance scans is the "S" or "R" as the first letter of the file name. Therefore whenever scans with names starting with "S" or "R" are read a dialog box will ask you to confirm whether these are in fact pseudo-absorbance data.

- The important functional difference between pseudo-absorbance data and true absorbance data is that pseudo-absorbance data exhibit systematic shifts (radially-invariant noise or RIN) somewhat analogous to "*jitter*" in interference scans. Thus when the program recognizes pseudo-absorbance data it turns on **jitter correction (Section 3.2.2.2.1)** by default.

6.3 About the fitting algorithm

This program uses a modified Gauss-Newton iterative least-squares fitting routine, essentially corresponding to the "preferred method" as described by Johnson & Faunt (1992) [**Reference 2 (Section 7.2)**].

The user may select whether all data points are equally weighted or to apply either theoretical or experimental weights. That choice is made within the **Select fitting model page (Section 3.2.3)**.

There are three modifications to the standard Gauss-Newton method, all of which occur at the end of each iteration:

1. Each new parameter is tested to be sure it falls within defined upper and lower bounds. If that parameter would exceed the bounded limit, it is set at the limiting value. When this occurs a yellow warning box appears next to that parameter on the **Fitting parameters monitor tab (Section 3.2.5.6)** of the **Perform least-squares fit page (Section 3.2.5)**.
2. An **increment reduction factor (Section 3.2.5.6.1)** is applied to insure that the new set of parameters actually gives a lower variance and to improve the rate of convergence. The current value of the increment reduction factor is also displayed on the **Fitting parameters monitor (Section 3.2.5.6)**.

3. If any **constraints (Section 3.2.3.7)** have been set on the fitting parameters those constrained parameters are updated to reflect the new values of the parameters to which they are linked.

6.4 About the meniscus wizard

1. For **absorbance scans** the program logic first locates a starting meniscus position by finding the highest OD value within the leftmost 20% of the last scan being used in the analysis (trying to find the top of the positive spike at the meniscus). It then examines a region within 0.02 cm of that initial position to find the top of the positive spike for each of the scans currently in use, and sets the meniscus at the arithmetic mean of those positions (rounded to the nearest 0.0001 cm after averaging).
    - This usually works well, since conventionally for absorbance scans the meniscus position is considered to at the top of the positive spike. When the sample OD is quite high the plateau is sometimes above the meniscus spike, and this algorithm may fail to find the correct meniscus position in such cases.
    - If you manually drag the meniscus marker you will then be asked whether you want the wizard to re-calculate the mean position of the spike. This is helpful when the scans contain two positive spikes, and the wizard has identified the wrong one as the true meniscus.
2. For **interference scans** there is usually a slight negative dip at the sample meniscus. Thus the logic positions the meniscus at the minimum fringe reading within the first 1/4 of the scan.
3. For **fluorescence scans** there is not yet sufficient data to discern a pattern to allow automatic location of the meniscus. Therefore at this time the program simply puts the meniscus at 6.0 cm (or whatever is the first data point if the scan starts to the right of 6.0 cm).

6.5 About the jitter wizard

To make the *jitter* correction automatically, the program tests two general regions, using spans of 20 data points:

1. a region in the air-air space
    - if the scans start to the right of 5.83 cm, the first 20 data points in each scan are used
    - otherwise a region started at 5.83 cm is tested; if the scan-to-scan variance is high at 5.83 cm (probably because the reference meniscus is to the left of 5.83 cm) the wizard compares regions starting at 5.82, 5.81, and 5.80 cm and picks the best one (lowest variance).
2. the region to the right of the meniscus starting at the left green "good" data marker (a region that will be constant if the meniscus is fully cleared at the time of analysis)
    - the initial trial region begins 0.007 cm to the right of the green data marker
    - four additional regions each 20 points further to the right are also examined to see which of these five possible regions is best (if the boundary is well clear of the meniscus generally a position further to the right, closer to the boundary, works best)

The wizard examines the deviations within each scan over these regions and picks the one where the variance is lower (after removing the average Y value for that scan over that region).

6.6 About the bootstrap method

In the 'bootstrap with replacement' (**reference 9 (Section 7.9)**) method (usually shortened to just 'bootstrap

method') for evaluating confidence limits data points are selected randomly from the original fitted data set to make a new data set with the same number of points. The random selection means that some of the original points will be selected more than once (and obviously others not at all). That is, each bootstrap data set is really a subset of the original data (but some points are repeated).

This new data set is then fitted, and the new fitted parameters are stored in an array. This whole process is repeated *N* times, and then for each fitted parameter the array of *N* values is used to evaluate the confidence limits.

Note that in the implementation used here some random noise is added to the best-fit parameters before each round so that each bootstrap round begins from slightly different starting values (to avoid biasing the results).

What are the advantages and drawbacks of this approach?

The principal advantage of the bootstrap method is that it makes no assumptions about the magnitude or randomness of the noise for the data being fitted. Thus for example if the *residual*s for the original best fit show a systematic pattern, that pattern is retained by the data subsets created during the bootstrap rounds.

A further advantage over the **Monte Carlo method (Section 6.7)** is that in real experiments the noise level of the raw data is often not constant across a scan or for scans at different times in the run. That is particularly true for fluorescence scans, where the noise magnitude is usually substantially higher in the plateau regions than in the low concentration regions to the left of the boundary. That same phenomenon also exists for absorbance scans, but the change in the noise magnitude is smaller than for fluorescence scans.

6.7 About the Monte Carlo method

In the Monte Carlo method for evaluating confidence limits first a set of theoretical raw scans is generated based on the number of species and the parameters from the best fit. Then random noise at a specified rms level is added to those theoretical raw scans, and a new data set is created (covering the same data range as the original data).

This new data set is then fitted, and the new fitted parameters are stored in an array. This whole process is repeated *N* times, and then for each fitted parameter the array of *N* values is used to evaluate the confidence limits.

Note that in the implementation used here some random noise is added to the best-fit parameters before each round so that each Monte Carlo round begins from slightly different starting values (to avoid biasing the results).

HHow is the magnitude of the random noise to be added determined?

The default value for the r.m.s. noise to be added equals the r.m.s. *residual* from the last fit. That is, the assumption is that all deviations are due to random noise in the data (there is no systematic pattern to the residuals), and that noise is the same for all radial regions and all scans in the run. For simulated data this value should be changed, if needed, to the actual noise level added to that simulation.

What are the advantages and drawbacks of this approach?

The principal advantage of the Monte Carlo method is that it evaluates the confidence limits based on noise in the raw scans (and noise that is truly random).

Its main drawbacks are:

1. we don't usually know the true noise level of the raw scans
2. the noise level may not be constant across each scan, or from scan to scan
    - in particular note that for fluorescence scans the noise magnitude usually is substantially lower in the region to the left of the boundary than in the plateau region---for that reason the **bootstrap method (Section 6.6)** may be more appropriate for fluorescence data

- the noise in absorbance scans is also usually significantly lower in the low OD regions to the left of the boundary than in the plateau region

3. we generally don't know whether the experimental noise is truly random (except when we are fitting simulations)

7 Bibliography

7.1 reference 1 - Philo (1997)

Philo, J. S. (1997). An improved function for fitting sedimentation velocity data for low-molecular-weight solutes. Biophys. J. 72, 435-444.

(This article can also be downloaded as a PDF file from the **program download page (http://www.jphilo.mailway.com/download.htm)** or directly **using this link (http://www.jphilo.mailway.com/sveddoc2.pdf)**.)

7.2 reference 2 - Johnson & Faunt (1992)

Johnson, M.L. And L.M. Faunt (1992). Parameter estimation by least-squares methods. *Methods Enzymol.* **210**, 1-37.

7.3 reference 3 - Stafford (1997)

Equation 2.127 from Fujita H. (1975), Foundations of Ultracentrifugal Analysis. (New York: John Wiley & Sons)

7.4 reference 4 - Durschlag (1986)

H. Durchschlag (1986). In *Thermodynamic Data for Biochemistry and Biotechnology*, H.-J. Hinz, ed., Springer-Verlag, New York, Chapter 3, p. 45.

7.5 reference 5 - Behlke & Ristau (2002)

The **Behlke function with base** corresponds to Equation 28 in Behlke, J. and Ristau, O. (2002). A new approximate whole boundary solution of the Lamm differential equation for the analysis of sedimentation velocity experiments. *Biophys. Chem.* 95, 59-68.

The **Behlke boundary function** corresponds to only the first 3 of the 6 terms in Eq. 28 (the moving boundary terms).

7.6 reference 6 - Balbo et al. (2009)

Brown, P. H., Balbo, A., and Schuck, P. (2009). On the analysis of sedimentation velocity in the study of protein complexes. *Eur. Biophys. J.* 38, 1079-1099.

7.7 reference 7 - Stafford (1994)

Philo, J. S. (2011). Limiting the sedimentation coefficient for sedimentation velocity data analysis: Partial boundary modeling and *g(s*)* approaches revisited. *Anal. Biochem.* 412, 189-202.

A PDF version of this paper can be downloaded using **this link (http://www.jphilo.mailway.com/Anal_Biochem_2011_189_202.pdf)**. The supplementary material for this article can be downloaded using **this link (http://www.jphilo.mailway.com/Anal_Biochem_2011_supplementary_material.pdf)**.

7.8 reference 8 - Schuck & Demeler 1999

Schuck, P. and Demeler, B. (1999). Direct sedimentation analysis of interference optical data in analytical ultracentrifugation. *Biophys. J.* 76, 2288-2296.

7.9 reference 9 - Efron & Tibshirani (1986)

Efron, B. and Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science* 1, 54-77.

7.10 reference 10 - Garcia de la Torre & Bloomfield (1981)

Garcia de la Torre, J. and V. A. Bloomfield (1981). Hydrodynamic properties of complex, rigid, biological macromolecules: Theory and applications. *Q. Rev. Biophys.* **14**: 81-139.

7.11 reference 11 - Johnson & Straume (1994)

Johnson, M. L. and Straume, M. (1994). Comments on the analysis of sedimentation equilibrium experiments. In: *Modern analytical ultracentrifugation: acquisition and interpretation of data for biological and synthetic polymer systems*. T.M.Schuster and T.M.Laue, eds. Birkhäuser, Boston, pp. 37-65.

7.12 reference 12 - Philo (2011)

Philo, J. S. (2011). Limiting the sedimentation coefficient for sedimentation velocity data analysis: Partial boundary modeling and *g(s*)* approaches revisited. *Anal. Biochem.* 412, 189-202.

7.13 reference 13- Arthur et al. (2009)

Arthur, K. K., Gabrielson, J. P., Kendrick, B. S., and Stoner, M. R. (2009). Detection of protein aggregates by

sedimentation velocity analytical ultracentrifugation (SV-AUC): Sources of variability and their relative importance. *J. Pharm. Sci.* 98, 3522-3539.

Index

1 iteration button, 148-149
About the bootstrap method, 206-207
About the fitting algorithm, 205-206
About the jitter wizard, 206
About the meniscus wizard, 206
About the Monte Carlo method, 207-208
About using radial intensity scans, 204-205
Absorbance scans, 77-80, 104
Add more scans for this sample button, 105
Adjust parameter upper/lower limits button, 144-145
Analysis page 1
- Load raw scans - Add more scans for this sample button, 105
- Load raw scans - Baseline scan usage, 107-109
- Load raw scans - Data loading options, 106-107
- Load raw scans - Delete selected scans button, 105-106
- Load raw scans - Load new scan files button, 104-105
- Load raw scans - Preselect data type and cell check box, 104
- Load raw scans - Replace scans for this sample button, 105
- Load raw scans - Run Claverie simulation to generate data button, 109-110
- Load raw scans - Sample description, 109
- Load raw scans - Scans currently loaded in memory list box, 106
- Load raw scans page, 102-104

Analysis page 2
- Set meniscus & data region - Cell base position, 118
- Set meniscus & data region - Fringe jump removal options, 115-116
- Set meniscus & data region - Good data begin X cm from meniscus, 117
- Set meniscus & data region - Good data end at radius =, 117-118
- Set meniscus & data region - Graph display options tab, 112-113
- Set meniscus & data region - Jitter removal options, 114-115
- Set meniscus & data region - Limit maximum buildup at cell base, 118-119
- Set meniscus & data region - Meniscus position, 116-117
- Set meniscus & data region - Systematic noise corrections options tab, 113-114
- Set meniscus & data region - Zoom meniscus button, 118
- Set meniscus & data region page, 110-112

Analysis page 3
- Select fitting model - Baseline fitting options, 131-132
- Select fitting model - Constrain M or s ratios among species, 133-136

Select fitting model - Convert to s(20w), 138-140
Select fitting model - Data options, 124-131
Select fitting model - Density and partial specific volumes, 137-138
Select fitting model - Fit diffusion or mass, 133
Select fitting model - Fitting function options, 123-124
Select fitting model - Import from SEDFIT c(s) button, 122-123
Select fitting model - Molar mass units, 136-137
Select fitting model - Multi-species options, 121-122
Select fitting model page, 119-121

Analysis page 4

Adjust parameter upper/lower limits button, 144-145
Constrained parameters, 143
Convergence control, 143-144
Fit control check boxes, 142-143
Parameter value text boxes, 142
Set/alter fitting parameters - Cancel changes button, 146
Set/alter fitting parameters page, 140-142
Slope correction fitting parameter, 145-146

Analysis page 5

Fitting parameters monitor tab, 149-151
Increment reduction factor, 151-152
Perform fit - 1 iteration button, 148-149
Perform fit - Cancel button, 149
Perform fit - Do fit button, 148
Perform fit - Simulate button, 149
Perform fit - Status, 149
Perform least-squares fit page, 146-148

Analysis page 6

Report fit results - Compute confidence intervals button, 156-159
Report fit results - Fit statistics, 160-161
Report fit results - Fitted parameters, 154-155
Report fit results - Fixed constrained & implicit parameters, 155
Report fit results - Graph bootstrap or Monte Carlo results , 161-164
Report fit results - High cross-correlation warning, 161
Report fit results - Print reports, 159-160
Report fit results - View parameter cross-correlations button, 155-156
Report fit results page, 152-154

Analysis page 7

Graph fit results page, 164-181

Analysis pages
Graph fit results, 164-181
Load raw scans page, 102-104
Perform least-squares fit, 146-148
Report fit results, 152-154
Select fitting model, 119-121
Set meniscus and data region, 110-112
Set/alter fitting parameters, 140-142
Analysis window, 96-98
Back button, 98
Bad data points, 189-190
Bad scans, 190-194
Baseline, 198-201
Baseline fitting options, 131-132
Baseline offset, 131-132
Baseline subtraction, 198-201
Behlke boundary function, 123-124
Behlke function with base, 123-124
Bibliography
Reference 1, 209
Reference 10, 210
Reference 11, 210
Reference 12, 210
Reference 13, 210-211
Reference 2, 209
Reference 3, 209
Reference 4, 209
Reference 5, 209
Reference 6, 209
Reference 7, 210
Reference 8, 210
Reference 9, 210
Bootstrap method, 156-159, 206-207
Buildup at cell base, 118-119
Buoyant mass, 136-137
Calculator, 85-87
Cancel button, 149
Cancel changes button, 146
Cell base, 117-118, 118, 118-119, 197, 197-198

Cell base position, 118
Cell number, 104
Citation, 203
Claverie simulation, 109-110
Clone, 83
Clone analysis to new window, 83
Confidence limits, 156-159
Constrained parameters, 143
Constraints, 133-136, 143
Convergence, 143-144
Convergence control, 143-144
Convert to s(20,w), 138-140
Copy data, 83-85
Correlation matrix, 155-156
Cross-correlation, 202
Data loading options, 106-107
Data options, 124-131
Data simulation, 77-80, 109-110
Data type, 104
Defaults, 89-93
Delete selected scans button, 105-106
Delta-c mode, 124-131
Density, 137-138
Despiking, 106-107
Diffusion, 85-87, 133
Do fit button, 148
Does SVEDBERG utilize multi-core processors?, 203
Edit data, 106, 189-190
Edit menu, 83-85
Error ellipse, 156-159
Export, 188-189
F statistics, 156-159
F statistics calculator, 87-89
FAQs
Citing this program, 203
Good fit, 202-203
High parameter cross-correlation, 202
Should I fit the cell base position?, 197-198
Should I fit the meniscus position?, 196-197

What is a 'sloping plateau'?, 201-202
Which systematic noise removal method?, 198-201
Why should I limit buildup at the cell base?, 197
Features for SEDFIT users, 3
File menu, 81-82
File open, 82-83
File save, 83
Fit cell base position, 197-198
Fit control check boxes, 142-143
Fit meniscus position, 196-197
Fit statistics, 160-161
Fit status, 149
Fitted parameters, 154-155
Fitting parameters monitor, 146-148, 149-151
Fitting parameters monitor tab, 149-151
Fixed, constrained & implicit parameters, 155
Fluorescence scans, 77-80, 104, 206-207, 207-208
Frequently-asked questions
Does SVEDBERG utilize multi-core processors?, 203
How should I cite this program?, 203
Should I fit the cell base position?, 197-198
Should I fit the meniscus position?, 196-197
What constitutes a good fit?, 202-203
What does high parameter cross-correlation mean?, 202
What is a 'sloping plateau'?, 201-202
Which method of systematic noise removal should I use?, 198-201
Why should I limit the buildup at the cell base?, 197
Fringe jump removal options, 113-114, 115-116
Fringe jumps, 113-114, 115-116
Getting Started
Features for SEDFIT users, 3
How to use this Help file, 2-3
Program overview and organization, 1-2
Good data end at radius =, 117-118
Good fit, 202-203
Graph customization, 185-188
Graph fit results page, 164-181
Graphical fit monitor, 146-148
Help menu, 94-96

Help on Help, 2-3
High cross-correlation warning, 161
How (and why) to equilibrate the rotor temperature, 183-184
How should I cite this program?, 203
How to
Avoid common mistakes, 182-183
Customize graphs, 185-188
Equilibrate the rotor temperature, 183-184
Export graph images or data, 188-189
Identify and remove rogue scans, 190-194
Minimize effects of sloping plateau, 194-195
Print graphs, 185
Remove bad data points, 189-190
Zoom with the mouse, 189
How to avoid common mistakes , 182-183
How to customize graphs, 185-188
How to export graph images or data to other programs, 188-189
How to identify and remove rogue scans, 190-194
How to minimize the effects of a sloping plateau, 194-195
How to remove bad data points from scan files, 189-190
How to use this Help file, 2-3
How to zoom with the mouse, 189
How to...
How (and why) to equilibrate the rotor temperature, 183-184
How to avoid common mistakes , 182-183
How to customize graphs, 185-188
How to export graph images or data to other programs, 188-189
How to identify and remove rogue scans, 190-194
How to minimize the effects of a sloping plateau, 194-195
How to print graphs, 185
How to remove bad data points from scan files, 189-190
How to zoom with the mouse, 189
Increment reduction factor, 151-152
Intensity scans, 204-205
Interference scans, 77-80, 104, 113-114
Jitter, 114-115, 198-201
Jitter removal options, 113-114, 114-115
Jitter wizard, 206
Lamm equation solutions, 123-124

Limit buildup at cell base, 118-119
Load every Nth scan, 106-107
Load new scan files button, 104-105
Load raw scans - Add more scans for this sample button, 105
Load raw scans - Baseline scan usage, 107-109
Load raw scans - Data loading options, 106-107
Load raw scans - Delete selected scans button, 105-106
Load raw scans - Load new scan files button, 104-105
Load raw scans - Preselect data type and cell check box, 104
Load raw scans - Replace scans for this sample button, 105
Load raw scans - Run Claverie simulation to generate data button, 109-110
Load raw scans - Sample description, 109
Load raw scans - Scans currently loaded in memory list box, 106
Load raw scans page, 102-104
Log page, 98-100
Mass/diffusion calculator, 85-87
Maximum buildup at cell base, 118-119
Meniscus, 116-117, 117, 196-197
Meniscus clearance, 117
Meniscus position, 116-117, 196-197
Meniscus wizard, 206
Menus
Edit menu, 83-85
File menu, 81-82
Help menu, 94-96
Tools menu, 85
View menu, 85
Window menu, 93-94
MFM function, 123-124
Mistakes, 182-183
Molecular mass, 85-87, 133
Molecular mass units, 136-137
Monte Carlo method, 156-159, 207-208
Multi-core processors, 203
Navigation tree, 100-102
Next and Back buttons, 98
Next button, 98
Number of species, 121-122
Offset, 198-201

Open, 82-83
Open saved analysis..., 82-83
Options dialog, 89-93
Out-of-bounds, 144-145
Overlay plot, 164-181
Overview, 1-2
Parallel computation, 203
Parameter bounds, 144-145
Parameter confidence limits, 156-159
Parameter constraints, 133-136
Parameter cross-correlation, 155-156, 161
Parameter errors, 156-159
Parameter limits, 144-145
Parameter precision, 143-144
Parameter value text boxes, 142
Partial boundary modeling, 124-131
Partial specific volume, 137-138
PBM, 124-131
Perform fit - 1 iteration button, 148-149
Perform fit - Cancel button, 149
Perform fit - Do fit button, 148
Perform fit - Simulate button, 149
Perform fit - Status, 149
Perform least-squares fit page, 146-148
Pitfalls, 182-183
Preferences, 89-93
Preselect, 104
Print graph, 188-189
Print graphs, 185
Print reports, 159-160
Program citation, 203
Program organization, 1-2
Program overview and organization, 1-2
Pseudo-absorbance, 113-114, 204-205
radially-invariant noise, 131-132, 198-201
Reference 1, 209
Reference 10, 210
Reference 11, 210
Reference 12, 210

Reference 13, 210-211
Reference 2, 209
Reference 3, 209
Reference 4, 209
Reference 5, 209
Reference 6, 209
Reference 7, 210
Reference 8, 210
Reference 9, 210
Reference information
About the bootstrap method, 206-207
About the fitting algorithm, 205-206
About the jitter wizard, 206
About the meniscus wizard, 206
About the Monte Carlo method, 207-208
About using radial intensity scans, 204-205
Symbols used, 204
Replace scans for this sample button, 105
Report fit results - Compute confidence intervals button, 156-159
Report fit results - Fit statistics, 160-161
Report fit results - Fitted parameters, 154-155
Report fit results - Fixed, constrained, & implicit parameters, 155
Report fit results - Graph bootstrap or Monte Carlo results , 161-164
Report fit results - High cross-correlation warning, 161
Report fit results - Print reports, 159-160
Report fit results - View parameter cross-correlations button, 155-156
Report fit results page, 152-154
Reports, 159-160
Residual plot, 164-181
RIN, 113-114, 114-115, 131-132, 198-201
Rogue scans, 190-194
s(20,w), 138-140
Sample description, 109
Save, 83, 83
Save analysis, 83
Save analysis as, 83
Save analysis as..., 83
Scan differencing, 124-131
Scans currently loaded list box, 106

SEDFIT-related features, 3, 56-77, 104-105, 122-123
Select fitting model - Baseline fitting options, 131-132
Select fitting model - Constrain M or s ratios among species, 133-136
Select fitting model - Convert to s(20,w), 138-140
Select fitting model - Data options, 124-131
Select fitting model - Density and partial specific volumes, 137-138
Select fitting model - Fit diffusion or mass, 133
Select fitting model - Fitting function options, 123-124
Select fitting model - Import from SEDFIT c(s) button, 122-123
Select fitting model - Molar mass units, 136-137
Select fitting model - Multi-species options, 121-122
Select fitting model page, 119-121
Set defaults, 89-93
Set meniscus & data region - Cell base position, 118
Set meniscus & data region - Fringe jump removal options, 115-116
Set meniscus & data region - Good data begin X cm from meniscus, 117
Set meniscus & data region - Good data end at radius =, 117-118
Set meniscus & data region - Graph display options tab, 112-113
Set meniscus & data region - Jitter removal options, 114-115
Set meniscus & data region - Limit maximum buildup at cell base, 118-119
Set meniscus & data region - Meniscus position, 116-117
Set meniscus & data region - Systematic noise corrections options tab, 113-114
Set meniscus & data region - Zoom meniscus button, 118
Set meniscus & data region page, 110-112
Set/alter fitting parameters - Cancel changes button, 146
Set/alter fitting parameters page, 140-142
Should I fit the cell base position?, 197-198
Should I fit the meniscus position?, 196-197
Simulate button, 149
Simulation, 77-80, 109-110
Slope correction fitting parameter, 145-146
Sloping plateau, 145-146, 201-202
Solvent density, 137-138
Species plot, 164-181
Statistics, 160-161
Summary plot, 164-181
Symbols used, 204
Synthetic boundary cells, 123-124
Synthetic boundary function, 123-124

Temperature equilibration, 183-184
time-invariant noise, 131-132, 198-201
TIN, 131-132, 198-201
Toolbar, 96
Tools menu, 85
True mass, 136-137
Tutorial: Multi-species fits, 34-55
Tutorial: Simulating experiments with the Claverie simulator, 77-80
Tutorial: Single species fit, 4-34
Tutorial: Using a c(s) distribution with SVEDBERG, 56-77
Tutorials
- Claverie simulation, 77-80
- Multi-species fits, 34-55
- Single species fit, 4-34
- Tutorial: Multi-species fits, 34-55
- Tutorial: Simulating experiments with the Claverie simulator, 77-80
- Tutorial: Single species fit, 4-34
- Tutorial: Using a c(s) distribution with SVEDBERG, 56-77

User interface elements
- Analysis window, 96-98
- Clone analysis to new window, 83
- Edit menu, 83-85
- F statistics calculator, 87-89
- File menu, 81-82
- Help menu, 94-96
- Log page, 98-100
- Mass/diffusion calculator, 85-87
- Navigation tree, 100-102
- Next and Back buttons, 98
- Open saved analysis..., 82-83
- Options dialog, 89-93
- Save analysis, 83
- Save analysis as..., 83
- Toolbar, 96
- Tools menu, 85
- User prompt area, 100
- View menu, 85
- Windows menu, 93-94

User prompt area, 100

Vbar, 137-138
View menu, 85
View parameter cross-correlations, 155-156
What constitutes a good fit?, 202-203
What does high parameter cross-correlation mean?, 202
What is a 'sloping plateau'?, 201-202
Which method of systematic noise removal should I use?, 198-201
Why should I limit the buildup at the cell base?, 197
Windows menu, 93-94
Zoom, 189
Zoom meniscus button, 118